Random Partitioning Problems Involving Poisson Point Processes On The Interval

Thierry HUILLET
Laboratoire de Physique Théorique et Modélisation,
CNRS-UMR 8089 et Université de Cergy-Pontoise,
2 Avenue Adolphe Chauvin, 95032, Cergy-Pontoise, FRANCE.
E-mail: Thierry.Huillet@ptm.u-cergy.fr

Abstract

Suppose some random resource (energy, mass or space) $\chi \geq 0$ is to be shared at random between (possibly infinitely many) species (atoms or fragments). Assume $\mathbb{E}\chi = \theta < \infty$ and suppose the amount of the individual share is necessarily bounded from above by 1. This random partitioning model can naturally be identified with the study of infinitely divisible random variables with Lévy measure concentrated on the interval. Special emphasis is put on these special partitioning models in the Poisson-Kingman class. The masses attached to the atoms of such partitions are sorted in decreasing order. Considering nearest-neighbors spacings yields a partition of unity which also deserves special interest. For such partition models, various statistical questions are addressed among which: correlation structure, cumulative energy of the first K largest items, partition function, threshold and covering statistics, weighted partition, Rényi's, typical and size-biased fragments size. Several physical images are supplied.

When the unbounded Lévy measure of χ is $\theta x^{-1} \cdot \mathbf{I}(x \in (0,1))\, dx$, the spacings partition has Griffiths-Engen-McCloskey or GEM(θ) distribution and χ follows Dickman distribution. The induced partition models have many remarkable peculiarities which are outlined.

The case with finitely many (Poisson) fragments in the partition law is also briefly addressed. Here, the Lévy measure is bounded.

KEYWORDS: Random Partitions, Divisibility, Poisson Point Process on the Interval.

AMS 1991: Primary 60G57, 62E17, Secondary: 60K99, 62E15, 62E20

1 Introduction

Random division models of a population into a (possibly large) number n of species, fragments or valleys with random weights or sizes have received considerable attention in various domains of applications.

In disordered systems Physics, it was first recognized as an important issue in [1], as a result of phase space (in iterated maps or spin glasses models at thermal equilibrium) being typically broken into many valleys, or attraction basins, each with random weight. Problems involving the breakdown or splitting of some item into random component parts or fragments, also appear in many other fields of interest: for example the composition of rocks into component compounds in Geology (splitting of mineral grains or pebbles), the composition of biological populations into species or the random allocation of memory in Computer Sciences, but also models for gene frequencies in population genetics and biological diversity.

All these applications deal with randomly broken objects and random splitting (see also [2] pages 25 and 30 for further motivations). Considering the random weights of the various species must sum to one, by normalization, the typical phase-space of these models is the interval of unit length, randomly split in such a way that the fragments' masses, sizes or energies must sum to one. The random structure of the population is then characterized by the ranked sequence of fragments' weights or sizes. This was observed in [3] (in the large n thermodynamic limit i.e. with a denumerable number of fragments).

There are of course many ways to break the interval at random into n (possibly infinitely many) pieces and so one needs to be more specific. This manuscript is precisely devoted to the study of some remarkable partition laws of the interval which arise from the partitioning problem of some random variable.

More precisely, in Section 2, we shall first focus on the simplest "fair" statistical model for splitting the interval into a finite number n of fragments. It essentially relies on normalization of a sequence of random variables by its sum. In more details, let $n > 1$ be a given integer. With $(S_k; k \geq 1)$ independent and identically distributed positive random variables, consider the partial random walk sum $\chi_n := S_1 + .. + S_n$. Then $S_1, ..., S_n$ constitute a simple random partition of χ_n. Normalizing, define $\varsigma_k = S_k/\chi_n$, $k = 1, .., n$. Then $(\varsigma_1, ..., \varsigma_n)$ constitutes a random partition of unity. In this model, there are $1 < n < \infty$ fragments with exchangeable random sizes $(\varsigma_1, ..., \varsigma_n)$ summing up to 1. We shall focus on the special case where S_1 has gamma(α) distribution, with $\alpha > 0$. In this case, $(\varsigma_1, ..., \varsigma_n)$ has Dirichlet $\mathrm{D}_n(\alpha)$ distribution. Let $\left(\varsigma_{(1)}, ..., \varsigma_{(n)}\right)$ be the ranked version of $(\varsigma_1, ..., \varsigma_n)$, with $\varsigma_{(1)} > ... > \varsigma_{(n)}$. Passing to the weak limit $n \uparrow \infty$, $\alpha \downarrow 0$ with $n\alpha = \theta$, one gets the ranked Poisson-Dirichlet partition model PD(θ). It may also be obtained from the normalization process of the jumps of a Moran subordinator, resulting in a random discrete distribution on the infinite simplex; see [4]. We shall recall some of its remarkable properties. The PD model exhibits many fundamental invariance properties. For a review of these results and applications to Computer Science, Combinatorial Structures, Physics, Biology.., see [5] and the references therein for example; this model and related ones are also fundamental in Probability Theory; see [6], [7], [8] and [9].

Several (not exclusively) interesting partitioning models of the interval are based

on such normalizing process in the literature. In the sequel, we shall study different types of partitioning models rather based on nearest-neighbor spacings.

More specifically, in Section 3, we shall indeed discuss the following closely related partitioning model. Let $\chi > 0$ be an infinitely divisible random variable with Lévy measure concentrated on $(0, 1)$ with total mass ∞ (the so-called unbounded case). Assume $\mathbb{E}\chi = \theta < \infty$. Then the partition $\chi \stackrel{d}{=} \sum_{k\geq 1} \xi_{(k)}$ is obtained from the constitutive ordered jumps $\xi_{(1)} > ... > \xi_{(k)} > ...$ of χ. The system $\left(\xi_{(k)}; k \geq 1\right)$ constitutes a Poisson point process on $(0, 1)$. Special emphasis is put on these partitioning models of the Poisson-Kingman type in Section 3.1. Their specificity is that each fragment in the decomposition of χ has size physically bounded from above by 1. Several statistical questions arising in this partitioning context are then discussed among which: fragments correlation structure, cumulative sum of the K largest items, partition function, threshold statistics, filtered partition, typical and size-biased picked fragments size. All these statistical questions are of concrete interest.

In Section 3.2, the following related partition is also considered: let $\widetilde{\xi}_k = \xi_{(k-1)} - \xi_{(k)}$, $k \geq 1$, (with $\xi_{(0)} := 1$) stand for spacings between consecutively ordered $\xi_{(k)}$s. Then, $\sum_{k\geq 1} \widetilde{\xi}_k = 1$ and $\left(\widetilde{\xi}_k, k \geq 1\right)$ constitutes an alternative random partition of unity. In sharp contrast with limiting partitioning models of Section 2, its construction does not involve any normalization procedure. Its specificity rather is a consequence of the Lévy measure for jumps of χ to be concentrated on $(0, 1)$ leading to $(0, 1)$ $-$valued $\xi_{(k)}$s. Similar statistical questions arising in this partition context are also addressed.

In Section 4, a remarkable special case of the partitioning models developed in Section 3 is studied in some detail. It corresponds to the following particular model: assume the unbounded Lévy measure of χ takes the particular form: $\theta/x \cdot \mathbf{I}_{x\in(0,1)}dx$. Then, the random variable χ has Dickman distribution. The induced partitioning models have many remarkable peculiarities which are outlined throughout. In particular, the spacings partition has Griffiths-Engen-McCloskey or GEM(θ) distribution whose ordered version is Poisson-Dirichlet partition. The PD(θ) was obtained in Section 2 from a very different construction based on normalization.

Finally, in Section 5, we assume that the Lévy measure of χ is now with finite total mass (the bounded case). In this case, we are led to random partitions of χ or of unity into a finite Poissonian number of fragments. Some of their properties are briefly outlined.

2 Exchangeable Dirichlet Partition with Finitely Many Fragments and its Poisson-Dirichlet Limit

We start with recalling a standard construction of the Poisson-Dirichlet partition as a limiting partition from the exchangeable Dirichlet partition of unity.

Dirichlet partition

Suppose there are $\infty > n > 1$ fragments with random sizes, say $(\varsigma_1, ..., \varsigma_n)$, where $(\varsigma_1, ..., \varsigma_n)$ has exchangeable distribution, implying in particular that each ς_k,

$k=1,..,n$ all share the same distribution, say the one of $\varsigma \overset{d}{=} \varsigma_1$ (each item has statistically the same mass). We also assume that ς has a density $f_\varsigma(x) > 0$ on $(0,1)$ with total mass 1 and that $\sum_{k=1}^n \varsigma_k = 1$ (almost surely) which is a strict conservativeness property of the partition.

With $\alpha > 0$, we assume more specifically that $(\varsigma_1,..,\varsigma_n)$ is distributed according to the (exchangeable) Dirichlet-$\mathrm{D}_n(\alpha)$ density function on the simplex meaning

$$f(x_1,..,x_n) = \frac{\Gamma(n\alpha)}{\Gamma(\alpha)^n} \prod_{k=1}^n x_k^{\alpha-1} \cdot \delta_{\left(\sum_{k=1}^n x_k - 1\right)}. \tag{2.1}$$

In this case, $\varsigma_k \overset{d}{=} \varsigma$, $k=1,..,n$ and the individual fractions are all identically distributed. Their common density on the interval $(0,1)$ is given by

$$f_\varsigma(x) = \frac{\Gamma(n\alpha)}{\Gamma(\alpha)\Gamma((n-1)\alpha)} x^{\alpha-1}(1-x)^{(n-1)\alpha-1}.$$

This is the one of a beta$(\alpha,(n-1)\alpha)$ random variable, with moment function

$$\varphi_\varsigma(q) := \mathbb{E}(\varsigma^q) = \frac{\Gamma(n\alpha)}{\Gamma(n\alpha+q)} \frac{\Gamma(\alpha+q)}{\Gamma(\alpha)},\ q > -\alpha. \tag{2.2}$$

The case $\alpha = 1$ corresponds to the uniform partition into n fragments for which

$$\varphi_\varsigma(q) = \frac{(n-1)!}{(q+n-1)(q+n-2)...(q+1)},\ q > -1.$$

This remarkable partition model is in the larger class of those for which $\varsigma_k = S_k/(S_1 + .. + S_n)$ where the S_k, $k=1,..,n$ are independent and identically distributed (iid) positive random variables. Indeed, assuming $S_1 \overset{d}{\sim}$ gamma(α), the joint distribution of $(\varsigma_1,..,\varsigma_n)$ is well-known to be Dirichlet $\mathrm{D}_n(\alpha)$.

Poisson-Dirichlet partition and the Kingman limit

In such "equitable" Dirichlet model, consider the situation where $n \uparrow \infty$, $\alpha \downarrow 0$ while $n\alpha = \theta > 0$. Such an asymptotic was first considered by [10]. As noted by Kingman, $(\varsigma_1,..,\varsigma_n) \overset{d}{\sim} \mathrm{D}_n(\alpha)$ itself has no non-degenerate limit. However, considering the ranked version $\left(\varsigma_{(1)},..,\varsigma_{(n)}\right)$ with $\varsigma_{(1)} > .. > \varsigma_{(n)}$, one may check that in the Kingman limit, $\left(\varsigma_{(k)}, k=1,..,n\right)$ converges in law to a Poisson-Dirichlet distribution, say $\left(\varsigma_{(k)},\ k \geq 1\right) \overset{d}{\sim}$ PD(θ) with $\varsigma_{(1)} > .. > \varsigma_{(k)} > ...$ The size-biased permutation of $\left(\varsigma_{(k)},\ k \geq 1\right)$ is, say $(\varsigma_k,\ k \geq 1) \overset{d}{\sim}$ GEM(θ), the so-called Griffiths-Engen-McCloskey law (see [4], Chapter 9). For this partition of unity, the following Residual Allocation Model (or RAM) decomposition holds

$$\varsigma_k = \prod_{l=1}^{k-1} \overline{v}_l v_k,\ k \geq 1. \tag{2.3}$$

Here $(v_k, k \geq 1)$ are iid with common law $v_1 \overset{d}{\sim}$ beta$(1,\theta)$ and $\overline{v}_1 := 1 - v_1 \overset{d}{\sim}$ beta$(\theta,1)$. Note that $\varsigma_1 \succeq_{st} .. \succeq_{st} \varsigma_k \succeq_{st} ..$, and that $(\varsigma_k,\ k \geq 1)$ is invariant under size-biased

permutation.

As is well-known, the Poisson-Dirichlet partition can be understood as follows (see [4], Chapter 9 and [11]). Let $\chi > 0$ be some infinitely divisible random variable whose Lévy measure $\Pi(dx)$ is concentrated on $(0, \infty)$, with infinite total mass. Assume more specifically that $\Pi(dx) = \theta e^{-x}/x$, $x > 0$. This is the Lévy measure for jumps of a Moran (gamma) subordinator $(\chi_t; t \geq 0)$, with $\chi := \chi_1 \overset{d}{\sim}$ gamma(θ). Let $\left(\xi_{(k)}, k \geq 1\right)$ be the ranked constitutive jumps of χ with $\chi \overset{d}{=} \sum_{k\geq 1} \xi_{(k)}$ and $\xi_{(1)} > .. > \xi_{(k)} > ....$ Let $\left(\varsigma_{(k)} := \xi_{(k)}/\chi;\ k \geq 1\right)$ be the ranked normalized jumps, hence with $1 = \sum_{k\geq 1} \varsigma_{(k)}$. Then, $\left(\varsigma_{(k)};\ k \geq 1\right)$ has Poisson-Dirichlet PD(θ) distribution, independent of $\chi = \chi_1$. This interpretation of $\left(\varsigma_{(k)};\ k \geq 1\right)$ in terms of normalized jumps of a gamma(θ)-distributed random variable is the limiting manifestation of the fact $\varsigma_k = S_k/\left(S_1 + .. + S_n\right)$, $k = 1, .., n$ (with S_k iid positive gamma(α)-distributed random variables), characterizing the Dirichlet $\mathrm{D}_n(\alpha)$ model.

3 Partitioning Constructions Based on Integrable Infinitely Divisible Random Variables with Lévy Measure Concentrated on $(0, 1)$: the Unbounded Case

Suppose some random resource (or energy, mass or amount of space) $\chi \geq 0$ is to be shared at random between (possibly infinitely many) species (atoms, fragments), the amount of the individual share being necessarily bounded by one. This random partitioning model can nicely be handled from infinitely divisible (ID) random variables with Lévy measure concentrated on $(0, 1)$ [See [12] and [13], for general monographs on infinite-divisibility].

If the physical interpretation of χ is energy (as in earthquake magnitude data with χ interpreting as the cumulative energy releases on Earth over some period of time), our construction is, to some extent, related to the Random Energy Model of Derrida (see [1] and [3] and its "Poissonian" reformulation by [14] and [15]). The partitioning nature of this problem is indeed well-known. In insurance models, the individual share can represent the amount of a particular claim resulting from some damage. In population genetics, it could interpret as species abundance in a large population. In a partition of mass problem, the share attributed to each of the constitutive element of the partition is generally called the fragment size or mass. In an economical context, the individuals share is their asset. In any case, the peculiarity of our model is that the individual share of the constitutive atoms of the partition are all physically necessarily bounded above.

3.1 Random Partition of "Energy": The Model

Let $\chi \geq 0$ be an infinitely divisible random variable with Lévy measure for jumps $\Pi(dx)$ supported by $(0, 1)$. Hence, with $\lambda \in \mathbb{R}$, let

$$\mathbb{E}e^{-\lambda\chi} = \exp\left\{-\int_0^1 \left(1 - e^{-\lambda x}\right)\Pi(dx)\right\} \tag{3.1}$$

be the entire analytic Laplace-Stieltjes Transform (LST) of χs law. We shall assume that $\mathbb{E}\chi = \theta < \infty$ so that $0 < \int_0^1 x\Pi(dx) = \theta < \infty$. We shall also assume that Π has a (continuous) density, say π. In this case, the density f_χ of χ exists and is easily

seen to solve the functional equation

$$xf_{\chi}(x) = \int_0^{x\wedge 1} f_{\chi}(x-z)\, z\pi(z)\, dz. \tag{3.2}$$

As is-well-known, the random variable χ is naturally associated to $(\chi_t, t \geq 0)$ which is a process with stationary independent increments. The process $(\chi_t, t \geq 0)$ is a subordinator with no drift and non-negative jumps restricted to $(0,1)$ and $\chi = \chi_1$.

Let $\overline{\Pi}(x) := \int_x^1 \Pi(dz)$. Two cases arise, depending on whether $\overline{\Pi}(0) = \infty$ (the unbounded case) or $\overline{\Pi}(0) < \infty$ (the bounded case). In this Section, we shall first address statistical issues arising in the unbounded partitioning model.

Random Partition of Energy χ (Unbounded Case): First Properties

Here $\overline{\Pi}(0) = \infty$, where $\overline{\Pi}(x) = \int_x^1 \Pi(dz)$ is the tail of the Lévy measure. In this case, the total mass of Π is infinite and $\chi > 0$. Plainly, we have

$$\chi \overset{d}{=} \sum_{k\geq 1} \overline{\Pi}^{-1}(S_k) \tag{3.3}$$

where $(S_k; k \geq 1)$ are points of a homogeneous Poisson point process (PPP) on the half-line (with $T_k := S_k - S_{k-1}$ iid and exp(1) distributed) and $\overline{\Pi}^{-1}$ the decreasing inverse of $\overline{\Pi}$.

This decomposition constitutes a random partition of the random variable χ in terms of the (infinitely many) ranked constitutive jumps of χ_1, all bounded by 1. Let $\xi_{(k)} := \overline{\Pi}^{-1}(S_k)$, $k \geq 1$, be such $(0,1)-$valued jumps arranged in decreasing order $\xi_{(1)} > .. > \xi_{(k)} > ...$ They constitute a PPP on $(0,1)$ with intensity Π, satisfying $\sum_{k\geq 1} \mathbb{E}\left(\xi_{(k)}\right) = \theta$. In the decomposition of χ model, with $\chi \overset{d}{=} \sum_{k\geq 1} \xi_{(k)}$, the random variable $\xi_{(k)}$ interprets as the kth fragment size.

When $\theta = 1$, the PPP system $\left(\xi_{(1)}, .., \xi_{(k)}, ..\right)$ is said to be conservative in average. If $\theta > 1$ ($\theta < 1$) we shall say that the system is excessive (defective).

First and Second Order Statistics: Correlation Structure

We start with supplying easy informations.

First, from the law of large numbers $\frac{1}{k}\overline{\Pi}\left(\xi_{(k)}\right) \to 1$ almost surely as $k \uparrow \infty$, supplying a useful information on the way $\xi_{(k)}$ goes to 0 when k grows.

Next, the one-dimensional distribution of $\xi_{(k)}$ is easily seen to be $\mathbb{P}\left(\xi_{(k)} \leq x\right) = \mathbb{P}\left(S_k > \overline{\Pi}(x)\right) = e^{-\overline{\Pi}(x)} \sum_{l=0}^{k-1} \frac{\overline{\Pi}(x)^l}{l!}$. In particular,

$$\mathbb{E}\left(\xi_{(k)}\right) = \frac{1}{\Gamma(k)} \int_0^\infty \overline{\Pi}^{-1}(s)\, s^{k-1} e^{-s} ds$$

$$\mathbb{E}\left(\xi_{(k)}^2\right) = \frac{1}{\Gamma(k)} \int_0^\infty \overline{\Pi}^{-1}(s)^2\, s^{k-1} e^{-s} ds$$

$$\mathbb{E}\left(\xi_{(k)}\xi_{(k+l)}\right) = \int_0^\infty \int_0^\infty \overline{\Pi}^{-1}(s_1)\, \overline{\Pi}^{-1}(s_1+s_2)\, \frac{s_1^{k-1}e^{-s_1} s_2^{l-1} e^{-s_2}}{\Gamma(k)\Gamma(l)} ds_1 ds_2$$

are the first, second moments of $\xi_{(k)}$, together with the joint second moment of $\xi_{(k)}$ and $\xi_{(k+l)}$, k and $l \geq 1$. From this last expression, there is no general stationarity property to be expected. As a result, with $w_k := \mathbb{E}\xi_{(k)}$, the second order quantity and its $(q_1, q_2)-$definition domain

$$C_l(q_1, q_2) := \mathbb{E}\left(\sum_{k\geq 1} w_k \xi_{(k)}^{q_1} \xi_{(k+l)}^{q_2}\right) \tag{3.4}$$

deserves some interest. It gives weight w_k to the kth contribution to the full pair-correlation function at distance l.

Example (Dickman)**:**

Assume $\Pi(dx) = \frac{\theta}{x}\mathbf{I}_{x\in(0,1)}dx$, then $\overline{\Pi}^{-1}(s) = \exp\{-s/\theta\}$. Then χs distribution is closely related to Dickman distribution (see below). In this case,

$$\mathbb{E}\left(\xi_{(k)}\right) = \left(\frac{\theta}{\theta+1}\right)^k, \sigma^2\left(\xi_{(k)}\right) = \left(\frac{\theta}{\theta+2}\right)^k - \left(\frac{\theta}{\theta+1}\right)^{2k},$$
$$\mathbb{E}\left(\xi_{(k)}\xi_{(k+l)}\right) = \left(\frac{\theta}{\theta+2}\right)^k \left(\frac{\theta}{\theta+1}\right)^l$$

are the mean, variance of $\xi_{(k)}$ and correlation of $\xi_{(k)}$, $\xi_{(k+l)}$. In this particular separable case, the covariance coefficient $cov\left(\xi_{(k)}, \xi_{(k+l)}\right) = \sigma^2\left(\xi_{(k)}\right)\left(\frac{\theta}{\theta+1}\right)^l$ and

$$\mathbb{E}\left(\sum_{k\geq 1}\xi_{(k)}\xi_{(k+l)}\right) = \frac{\theta}{2}\left(\frac{\theta}{\theta+1}\right)^l$$

has exponential decay with l. More generally, in this particular case, with definition domain $q_1 + q_2 > -\theta/(\theta+1)$ and $q_2 > -\theta$, we obtain

$$C_l(q_1, q_2) = \frac{\theta^2}{\theta + (\theta+1)(q_1+q_2)}\left(\frac{\theta}{\theta+q_2}\right)^l. \tag{3.5}$$

This particular Dickman-model exhibits many other remarkable properties; these will be emphasized in some detail in the sequel. □

Campbell Formula

A very useful formula in our context is Campbell formula. We first recall it and then show its usefulness in the computation of statistical variables of concrete interest in the partitioning problem under study.

Let $\lambda \geq 0$ and g be a measurable function such that $\int_0^1 \left(1 - e^{-\lambda g(x)}\right)\Pi(dx) < \infty$, then by Campbell formula (see [16])

$$\mathbb{E}\exp\left\{-\lambda\sum_{k\geq 1} g\left(\overline{\Pi}^{-1}(S_k)\right)\right\} = \exp\left\{-\int_0^1 \left(1 - e^{-\lambda g(x)}\right)\Pi(dx)\right\} \tag{3.6}$$

is the Laplace-Stieltjes transform (LST) of $\sum_{k\geq 1} g\left(\overline{\Pi}^{-1}(S_k)\right)$. In particular, its mean value is

$$\mathbb{E}\sum_{k\geq 1} g\left(\overline{\Pi}^{-1}(S_k)\right) = \int_0^1 g(x)\,\Pi(dx).$$

Let us draw some conclusions of these elementary facts.

Cumulative energy of the $K-$biggest and of the remaining events

Let $K \geq 1$. The random variable $\chi_K^- := \sum_{k>K} \xi_{(k)}$ represents the amount of total energy χ concentrated in the lowest energy levels (at rank $K+1$ and below). Let us consider the problem of computing its law. We clearly have $\chi_K^- \overset{d}{=} \sum_{k\geq 1} \overline{\Pi}^{-1}(S_K + S_k)$ where the PPP $(S_k; k \geq 1)$ is independent of $S_K \overset{d}{\sim} \text{gamma}(K)$. As a result, applying Campbell formula (3.6) with $g(x) = \overline{\Pi}^{-1}\left(S_K + \overline{\Pi}(x)\right)$, we obtain

$$\mathbb{E}\exp\left\{-\lambda\chi_K^-\right\} = \mathbb{E}\exp\left\{-\lambda\sum_{k\geq 1}\overline{\Pi}^{-1}\left(S_K + \overline{\Pi}\left(\xi_{(k)}\right)\right)\right\}$$

$$= \mathbb{E}\exp\left\{-\int_0^1\left(1 - e^{-\lambda\overline{\Pi}^{-1}\left(S_K + \overline{\Pi}(x)\right)}\right)\Pi(dx)\right\}, \tag{3.7}$$

where the last expectation is over S_K. Note that $\chi = \chi_K^+ + \chi_K^-$ where $\chi_K^+ := \sum_{k=1}^K \xi_{(k)}$ is the contribution of the K largest energy levels to χ.

This question is closely related to the following problem: let $x > 0$ be some threshold value. Define

$$K(x) := \inf\left(K \geq 1 : \chi_K^+ > x\right)$$

to be the first time the cumulated fragments size of $\left(\xi_{(k)}, k \geq 1\right)$ exceeds x. Then, $\mathbb{P}(K(x) > K) = \mathbb{P}\left(\chi_K^+ \leq x\right)$ and the distribution of $K(x)$ results from the one of χ_K^+. See [17] for similar considerations.

Example (Dickman)*:*
Assuming $\Pi(dx) = \frac{\theta}{x}\mathbf{I}_{x\in(0,1)}dx$, then $\overline{\Pi}^{-1}(s) = \exp\{-s/\theta\}$ and

$$\begin{aligned}\mathbb{E}\exp\left\{-\lambda\chi_K^-\right\} &= \mathbb{E}\exp\left\{-\int_0^1\left(1 - e^{-\lambda\overline{\Pi}^{-1}\left(S_K + \overline{\Pi}(x)\right)}\right)\Pi(dx)\right\} \\ &= \mathbb{E}\exp\left\{-\theta\int_0^1\left(1 - e^{-\lambda x e^{-\frac{S_K}{\theta}}}\right)\frac{1}{x}dx\right\}.\end{aligned}$$

This shows that, in this particular case,

$$\chi_K^- \overset{d}{=} R_K \cdot \chi \text{ and } \chi_K^+ \overset{d}{=} (1 - R_K)\cdot\chi$$

where $R_K := \exp\left\{-\frac{S_K}{\theta}\right\} \in (0,1)$ is log-gamma(K,θ) distributed, with $\mathbb{E}R_K^q = [\theta/(\theta+q)]^K$, independent of χ. When $K = \left[(\log_2(1+1/\theta))^{-1}\right]$, the average wealth $\mathbb{E}\chi_K^-$ is half the one of χ.

In this example, we shall show below that χ has a Dickman type distribution, see (4.5) below, resulting in an intricate distribution for χ_K^- and χ_K^+ and consequently of $K(x)$. $\Box$

Partition Function of χ

Partition functions of energy are interesting quantities. Taking in particular $g(x) = x^\beta$ in (3.6), the full Laplace-Stieltjes transform of $\sum_{k\geq 1} \xi_{(k)}^\beta$ is obtained as

$$\mathbb{E}\exp\left\{-\lambda \sum_{k\geq 1} \xi_{(k)}^\beta\right\} = \exp\left\{-\int_0^1 \left(1 - e^{-\lambda x^\beta}\right)\Pi(dx)\right\}. \tag{3.8}$$

Its mean value

$$\phi(\beta) := \mathbb{E}\sum_{k\geq 1} \xi_{(k)}^\beta = \int_0^1 x^\beta \Pi(dx) \tag{3.9}$$

is defined for values of $\beta > \beta_*$ for which $\int_0^1 x^\beta \Pi(dx) < \infty$, with

$$\beta_* := \sup(\beta : \phi(\beta) = \infty) \in [0, 1).$$

Note indeed that $\phi(0) = \infty$ and $\phi(1) = \theta$.

In this setup, the Lévy measure Π interprets as follows.

Let $N_+(x) := \sum_{k\geq 1} \mathbf{I}\left(\xi_{(k)} > x\right)$ be the random number of $\xi_{(k)}$ exceeding $x \in (0, 1)$. Then $\overline{\Pi}(x)$ is the expected value of this number. Indeed

$$\begin{aligned}
\overline{\Pi}(x) &= \sum_{k\geq 1} \mathbb{P}\left(\xi_{(k)} > x\right) = \sum_{k\geq 1} \mathbb{P}\left(S_k \leq \overline{\Pi}(x)\right) \\
&= e^{-\overline{\Pi}(x)} \sum_{k\geq 1}\sum_{l\geq k} \frac{\overline{\Pi}(x)^l}{l!} = e^{-\overline{\Pi}(x)} \sum_{l\geq 1} \frac{\overline{\Pi}(x)^l}{(l-1)!}.
\end{aligned}$$

The random variable $\sum_{k\geq 1} \xi_{(k)}^\beta$ is called the partition function of χ and Π its structural (or occupation) measure.

The Numbers of Atoms of χ above Cutoff ϵ and the Contribution to Total Mass of those Atoms above and below ϵ: Threshold Statistics

The above considerations naturally suggest the following problems of interest in Statistics.

- Upper-threshold Statistics.

If $\epsilon \in (0, 1)$ is some cutoff or threshold value, let $N_+(\epsilon)$ count the numbers of atoms of the partition of χ exceeding ϵ. If χ is the amount of some natural resource to be shared between infinitely many agents on the market, ϵ stands for the minimal individual wealth below which each agent should be considered as indigent (e.g. below the poverty line). If χ stands for "energy", ϵ could interpret as the level below which micro-events are undetectable by the currently available measuring devices (if

one thinks of a sequence of earthquakes magnitude data for example).

By Campbell formula

$$\mathbb{E}\exp\left\{-\lambda N_{+}\left(\epsilon\right)\right\} = \exp\left\{-\int_{0}^{1}\left(1-e^{-\lambda \mathbf{I}(x>\epsilon)}\right)\Pi\left(dx\right)\right\}$$

$$= \exp\left\{-\overline{\Pi}\left(\epsilon\right)\left(1-e^{-\lambda}\right)\right\} \tag{3.10}$$

is the full Laplace-Stieltjes transform of $N_{+}\left(\epsilon\right)$. This shows that $N_{+}\left(\epsilon\right)$ is in fact Poisson distributed with intensity $\overline{\Pi}\left(\epsilon\right)$. Recalling $\overline{\Pi}\left(\epsilon\right) \underset{\epsilon\downarrow 0}{\rightarrow} \infty$, the law of large numbers gives

$$N_{+}\left(\epsilon\right)/\overline{\Pi}\left(\epsilon\right) \overset{a.s.}{\rightarrow} 1,\ \epsilon\downarrow 0. \tag{3.11}$$

That $N_{+}\left(\epsilon\right)$ is Poisson distributed may be also checked as follows: we have $N_{+}\left(\epsilon\right) = \inf\left(k\geq 1 : \xi_{(k)}\leq\epsilon\right)-1$ and $\mathbb{P}\left(N_{+}\left(\epsilon\right)\geq k\right) = \mathbb{P}\left(\xi_{(k)}>\epsilon\right)$. This is also $\mathbb{P}\left(S_{k}\leq\overline{\Pi}\left(\epsilon\right)\right) = e^{-\overline{\Pi}(\epsilon)}\sum_{l\geq k}\frac{\overline{\Pi}(\epsilon)^{l}}{l!}$ and $N_{+}\left(\epsilon\right)$ is Poisson with intensity $\overline{\Pi}\left(\epsilon\right)$.

Remark (randomization of the cutoff):

A slightly more general problem is to consider the random variable $N_{+}\left(\epsilon_{1}\right) := \sum_{k\geq 1}\mathbf{I}\left(\xi_{(k)}>\epsilon_{k}\right)$ where $\left(\epsilon_{k},\, k\geq 1\right)$ are iid $\left(0,1\right)-$valued random variables, independent of $\left(\xi_{(k)},\, k\geq 1\right)$. In this model, the poverty threshold attached to each agent is assumed random but drawn from the same distribution and with mutual independence.

From Campbell formula, we obtain

$$\mathbb{E}\exp\left\{-\lambda N_{+}\left(\epsilon_{1}\right)\right\} = \exp\left\{-\mathbb{E}\overline{\Pi}\left(\epsilon_{1}\right)\left(1-e^{-\lambda}\right)\right\},$$

showing that $N_{+}\left(\epsilon_{1}\right)$ is Poisson distributed with intensity $\mathbb{E}\overline{\Pi}\left(\epsilon_{1}\right)$ if $\mathbb{E}\overline{\Pi}\left(\epsilon_{1}\right)<\infty$.

This is useful in the problem of random covering of $\left(\xi_{(k)},\, k\geq 1\right)$ by random intervals with sizes $\left(\epsilon_{k},\, k\geq 1\right)$. In particular, the covering probability is

$$\mathbb{P}\left(N_{+}\left(\epsilon_{1}\right)=0\right) = \exp\left\{-\mathbb{E}\overline{\Pi}\left(\epsilon_{1}\right)\right\}.$$

Assuming (Dickman): $\overline{\Pi}\left(\epsilon\right) = -\theta\log\epsilon$ and $\epsilon_{1}\overset{d}{\sim}$ Uniform$(0,1)$, the intensity reads $\mathbb{E}\overline{\Pi}\left(\epsilon_{1}\right) = -\theta\int_{0}^{1}\log\epsilon d\epsilon = \theta$. In this case, $N_{+}\left(\epsilon_{1}\right)$ simply is Poisson(θ) distributed. □

The contribution to total mass χ of those $\xi_{(k)}$ above $\epsilon\in(0,1)$, which is

$$\chi_{+}\left(\epsilon\right) := \sum_{k\geq 1}\xi_{(k)}\mathbf{I}\left(\xi_{(k)}>\epsilon\right) = \sum_{k=1}^{N_{+}(\epsilon)}\xi_{(k)},$$

is such that

$$\mathbb{E}\exp\left\{-\lambda\chi_{+}\left(\epsilon\right)\right\} = \exp\left\{-\int_{0}^{1}\left(1-e^{-\lambda x\mathbf{I}(x>\epsilon)}\right)\Pi\left(dx\right)\right\}$$

$$= \exp\left\{-\int_{\epsilon}^{1}\left(1-e^{-\lambda x}\right)\Pi\left(dx\right)\right\}.$$

This is the LST of an infinitely divisible (ID) random variable with Lévy measure Π concentrated on $(\epsilon, 1)$ for which clearly $\chi_+(\epsilon) \xrightarrow{d} \chi$ $(\epsilon \downarrow 0)$. This random variable is of the compound Poisson type since

$$\mathbb{E}\exp\{-\lambda\chi_+(\epsilon)\} = \exp\left\{-\overline{\Pi}(\epsilon)\left(1 - \int_\epsilon^1 e^{-\lambda x}\Pi(dx)/\overline{\Pi}(\epsilon)\right)\right\}. \tag{3.12}$$

Here indeed, $F_\epsilon(dx) := \Pi(dx)/\overline{\Pi}(\epsilon)$ is a probability distribution. Hence, with u_k, $k \geq 1$ an iid $(0,1)-$valued uniform sequence, P_ϵ a Poisson random variable with intensity $\overline{\Pi}(\epsilon)$, $\chi_+(\epsilon) = \sum_{k=1}^{P_\epsilon} \overline{F}_\epsilon^{-1}(u_k)$ belongs to the class of compound Poisson random variables. Stated differently

$$\chi_+(\epsilon) = \sum_{k=1}^{P_\epsilon} \overline{F}_\epsilon^{-1}\left(u_{(k),P_\epsilon}\right) \tag{3.13}$$

where $u_{(1),P_\epsilon} < .. < u_{(P_\epsilon),P_\epsilon}$ is obtained from uniform sample $u_1, .., u_{P_\epsilon}$ while ordering the constitutive terms. We note that $\chi_+(\epsilon)$ has an atom at $\chi_+(\epsilon) = 0$ with probability $e^{-\overline{\Pi}(\epsilon)}$. This partition is also $\chi_+(\epsilon) \overset{d}{=} \sum_{k=1}^{P_\epsilon} \overline{\Pi}^{-1}\left(\overline{\Pi}(\epsilon)\, u_{(k),P_\epsilon}\right)$ where, when $\epsilon \downarrow 0$, $P_\epsilon \overset{a.s.}{\to} \infty$ and $\left(\overline{\Pi}(\epsilon)\, u_{(1),P_\epsilon}, .., \overline{\Pi}(\epsilon)\, u_{(k),P_\epsilon}\right) \xrightarrow{d} (S_1, .., S_k, ..)$ a Poisson point process on $\mathbb{R}^+$. Thus the decomposition of $\chi_+(\epsilon)$ constitutes a weak Poisson-partition approximation to the one of χ.

Remarks:

(i) A slightly more general problem is to consider the random variable $\chi_+(\epsilon_1) := \sum_{k\geq 1} \xi_{(k)} \mathbf{I}\left(\xi_{(k)} > \epsilon_k\right)$ where $(\epsilon_k,\, k \geq 1)$ are iid $(0,1)-$valued random variables, independent of $\left(\xi_{(k)},\, k \geq 1\right)$. From Campbell formula, performing an integration by parts, with $F_{\epsilon_1}(\epsilon) = \mathbb{P}(\epsilon_1 \leq \epsilon)$, we obtain

$$\begin{aligned} \mathbb{E}\exp\{-\lambda\chi_+(\epsilon_1)\} &= \exp\left\{-\mathbb{E}\int_{\epsilon_1}^1 \left(1 - e^{-\lambda x}\right)\Pi(dx)\right\} \\ &= \exp\left\{-\int_0^1 \left(1 - e^{-\lambda x}\right) F_{\epsilon_1}(x)\,\Pi(dx)\right\} \end{aligned} \tag{3.14}$$

showing that, in general, $\chi_+(\epsilon_1)$ is an ID random variable with Lévy measure for jumps $F_{\epsilon_1}(x)\,\Pi(dx)$.

Note also that if $\chi_-(\epsilon_1) := \sum_{k\geq 1} \xi_{(k)} \mathbf{I}\left(\xi_{(k)} \leq \epsilon_k\right)$, clearly

$$\mathbb{E}\exp\{-\lambda\chi_-(\epsilon_1)\} = \exp\left\{-\int_0^1 \left(1 - e^{-\lambda x}\right) \overline{F}_{\epsilon_1}(x)\,\Pi(dx)\right\} \tag{3.15}$$

where $\overline{F}_{\epsilon_1}(x) := 1 - F_{\epsilon_1}(x)$.

(ii) Finally, in the random covering of $\left(\xi_{(k)},\, k \geq 1\right)$ by random intervals $(\epsilon_k,\, k \geq 1)$ context, the quantity $\chi_g := \sum_{k\geq 1} \left(\xi_{(k)} - \epsilon_k\right)_+$ interprets as the total gaps' length (in the economical context, it is the excess-wealth of the well-off agents). We obtain directly

$$\mathbb{E}\exp\left\{-\lambda \sum_{k\geq 1} \left(\xi_{(k)} - \epsilon_k\right)_+\right\} = \exp\left\{-\mathbb{E}\int_{\epsilon_1}^1 \left(1 - e^{-\lambda(x-\epsilon_1)}\right)\Pi(dx)\right\}$$

$$= \exp\left\{-\int_0^1 \left(1-e^{-\lambda z}\right)\mathbb{E}\Pi_{\epsilon_1}(dz)\right\}, \tag{3.16}$$

where $\Pi_{\epsilon_1}(dz)$ is the image measure of $\Pi(dx)$ by the application $x \to z = x - \epsilon_1 \in (0, 1-\epsilon_1)$. This is the LST of an ID random variable with Lévy measure for jumps $\mathbb{E}\Pi_{\epsilon_1}(dz)$. $\square$

Examples (Dickman)*:*
Assuming $\Pi(dx) = \frac{\theta}{x}\mathbf{I}_{x\in(0,1)}dx$, we have $\Pi_{\epsilon_1}(dz) = \frac{\theta}{z+\epsilon_1}\mathbf{I}_{z\in(0,1-\epsilon_1)}dz$.
• If in addition, ϵ_1 is uniformly distributed on $(0,1)$, we find explicitly

$$\mathbb{E}\Pi_{\epsilon_1}(dz) = dz\int_0^{1-z}\frac{\theta}{z+\epsilon}d\epsilon = -\theta\log z dz, z \in (0,1).$$

In the chosen example, the total gaps' length is an ID (rate θ compound Poisson) random variable with logarithmic density for jumps $-\log z\mathbf{I}_{z\in(0,1)}$ and $\mathbb{E}\chi_g/\mathbb{E}\chi = -\int_0^1 x\log x dx < 1$ is the average reduction factor.

• If ϵ_1 is not random, with $\epsilon_1 \overset{d}{\sim} \delta_{\epsilon_1-\epsilon}$, the total gaps' length is a (rate $-\theta\log\epsilon$) compound Poisson random variable with density for jumps: $\frac{-1}{(z+\epsilon)\log\epsilon}\mathbf{I}_{z\in(0,1-\epsilon)}$. In addition, one gets $\mathbb{E}\chi_g/\mathbb{E}\chi = 1-\epsilon+\epsilon\log\epsilon \to 1$ $(\epsilon \downarrow 0^+)$.

Incidentally, note that the lack of wealth of the poorest, which is $\sum_{k\geq 1}\left(\epsilon_k - \xi_{(k)}\right)_+$, diverges. $\square$

• Sub-threshold Statistics.

Similarly, let $N_-(\epsilon) := \sum_{k\geq 1}\mathbf{I}\left(\xi_{(k)} \leq \epsilon\right)$ count the random number of $\xi_{(k)}$ below cutoff $\epsilon \in (0,1)$, then $N_-(\epsilon) = \infty$ for all such ϵ. The contribution to total mass χ of those $\xi_{(k)}$ below $\epsilon \in (0,1)$, which is

$$\chi_-(\epsilon) := \sum_{k\geq 1}\xi_{(k)}\mathbf{I}\left(\xi_{(k)} \leq \epsilon\right) = \sum_{k>N_+(\epsilon)}\xi_{(k)}$$

is such that

$$\begin{aligned}\mathbb{E}\exp\{-\lambda\chi_-(\epsilon)\} &= \exp\left\{-\int_0^1\left(1-e^{-\lambda x\mathbf{I}(x\leq\epsilon)}\right)\Pi(dx)\right\}\\ &= \exp\left\{-\int_0^\epsilon\left(1-e^{-\lambda x}\right)\Pi(dx)\right\}.\end{aligned}$$

This is the LST of an ID random variable with Lévy measure Π concentrated on $(0,\epsilon)$, showing that $\chi_-(\epsilon)$ and $\chi_+(\epsilon)$ are independent with $\chi \overset{d}{=} \chi_-(\epsilon) + \chi_+(\epsilon)$. Furthermore, $\mathbb{E}\chi_-(\epsilon) = \int_0^\epsilon x\Pi(dx) \sim \epsilon^2\pi(\epsilon) \to_{\epsilon\downarrow 0} 0$ and the variance $\sigma^2[\chi_-(\epsilon)] \sim \epsilon^3\pi(\epsilon) \to_{\epsilon\downarrow 0} 0$. As a result,

$*$ If $\sigma[\chi_-(\epsilon)]/\mathbb{E}\chi_-(\epsilon) \underset{\epsilon\downarrow 0}{\to} 0$, one can check that the Central Limit Theorem holds

$$\frac{\chi_-(\epsilon) - \mathbb{E}\chi_-(\epsilon)}{\sigma[\chi_-(\epsilon)]} \overset{d}{\underset{\epsilon\downarrow 0}{\to}} \mathcal{N}(0,1). \tag{3.17}$$

$*$ If $\epsilon\pi(\epsilon) \to_{\epsilon\downarrow 0} a > 0$, then $\sigma[\chi_-(\epsilon)]/\mathbb{E}\chi_-(\epsilon) \underset{\epsilon\downarrow 0}{\to} 1/\sqrt{a}$ and we easily find

$$\frac{\chi_-(\epsilon)}{\mathbb{E}\chi_-(\epsilon)} \underset{\epsilon\downarrow 0}{\overset{d}{\to}} \chi_a \text{ with } \mathbb{E}e^{-\lambda\chi_a} = e^{-a\int_0^1\left(1-e^{-\lambda x/a}\right)dx/x}. \tag{3.18}$$

The limiting random variable χ_a is (mean 1) infinitely divisible with major interest. It will be studied in some detail in the sequel (see Subsection 4.1).

Although the atoms below the cutoff are infinitely many, their contribution to total mass always goes to 0 as the cutoff approaches 0.

Weighted Partitions (Modulation)

Let $(\mu_k; k \geq 1)$ be a sequence of iid non-negative random variables, independent of $\left(\xi_{(k)}; k \geq 1\right)$. We shall investigate some properties of the weighted or modulated random sequence $\left(\mu_k\xi_{(k)}; k \geq 1\right)$ as a new random transformed partition of $\chi_{\mu_1} := \sum_{k\geq 1}\mu_k\xi_{(k)}$. This question appears in the following problem: assume the events $\left(\xi_{(k)}; k \geq 1\right)$, summing up to χ, are each corrupted by some multiplicative independent noise $(\mu_k; k \geq 1)$; then the observed sequence of events becomes $\left(\mu_k\xi_{(k)}; k \geq 1\right)$ and the observed cumulative energy turns out to be χ_{μ_1}.

Let us first consider the quantity $\mathbb{E}\exp\left\{-\lambda\sum_{k\geq 1} g\left(\mu_k\xi_{(k)}\right)\right\}$. From Campbell formula, we have

$$\begin{aligned}\mathbb{E}\exp\left\{-\lambda\sum_{k\geq 1} g\left(\mu_k\xi_{(k)}\right)\right\} &= \mathbb{E}\left\{\prod_{k\geq 1}\mathbb{E}\left(\exp -\lambda g\left(\mu_k\xi_{(k)}\right) \mid \xi_{(k)}\right)\right\}\\ &= \exp\left\{-\mathbb{E}\int_0^1\left(1-e^{-\lambda g(\mu_1 x)}\right)\Pi(dx)\right\}\\ &= \exp\left\{-\mathbb{E}\int_0^{\mu_1}\left(1-e^{-\lambda g(z)}\right)\Pi_{\mu_1}(dz)\right\}\end{aligned} \tag{3.19}$$

with $\Pi_{\mu_1}(dz)$ the image measure of $\Pi(dx)$ by the application $x \to z = \mu_1 x$.

Examples (Dickman)*:*

We note the scale-invariance property $\Pi_{\mu_1}(dz) = \Pi(dz)$ when $\Pi(dx) = \frac{\theta}{x}dx$. Using an integration by parts and putting $\overline{F}_{\mu_1}(x) = \mathbb{P}(\mu_1 > x)$, this shows that in this particular case for Π only

$$\mathbb{E}\exp\left\{-\lambda\sum_{k\geq 1} g\left(\mu_k\xi_{(k)}\right)\right\} = e^{-\int_0^\infty\left(1-e^{-\lambda g(x)}\right)\overline{F}_{\mu_1}(x)\Pi(dx)}. \tag{3.20}$$

In particular, $\sum_{k\geq 1}\mu_k\xi_{(k)}$ is a positive ID random variable with no negative jumps whose induced Lévy measure for jumps is $\frac{\theta\overline{F}_{\mu_1}(x)}{x}dx$. For example

(i) p-thinning: if $\mu_1 \overset{d}{\sim}$ Bernoulli(p), the new Lévy measure is $\frac{p\theta}{x}\mathbf{I}_{x\in(0,1)}dx$.

(ii) uniform thinning: if $\mu_1 \overset{d}{\sim}$ Uniform$(0,1)$, the new transformed Lévy measure is $\frac{\theta(1-x)}{x}\mathbf{I}_{x\in(0,1)}dx$.

(iii) exponential scaling: if $\mu_1 \overset{d}{\sim} \exp(1)$, the new Lévy measure is $\frac{\theta}{x}e^{-x}\mathbf{I}_{x>0}dx$ which is the one of a gamma (or Moran) subordinator. Note that $\mu_1\xi_{(1)} \succeq_{st} .. \succeq_{st} \mu_k\xi_{(k)} \succeq_{st}$ ..and that, although the sequence $\left(\mu_k\xi_{(k)}; k \geq 1\right)$ is not strongly ordered by decreasing sizes, the constitutive terms sum up to a gamma(θ) $-$distributed random variable. This should not be confused with the other natural partition of $\chi \overset{d}{\sim}$ gamma(θ) given by

$$\chi \overset{d}{=} \sum_{k\geq 1} \varsigma_{(k)}, \text{ with } \varsigma_{(1)} > .. > \varsigma_{(k)} > ..$$

where $\varsigma_{(k)} = \overline{\Pi}^{-1}(S_k)$, $k \geq 1$ and $\overline{\Pi}(x) = \int_x^\infty \frac{\theta}{z} \exp\{-z\} dz$. This constitutes an example where two distinct sequences $\left(\mu_k\xi_{(k)}; k \geq 1\right)$ and $\left(\varsigma_{(k)}; k \geq 1\right)$ both share the same partition function. $\square$

Typical Fragment Size from $\left(\xi_{(k)}, k \geq 1\right)$

We can define the typical fragment size from $\left(\xi_{(k)}, k \geq 1\right)$ to be a $(0,1)$ $-$valued random variable, say ξ, with density $f_\xi(x)$, whose distribution function $F_\xi(x)$ is defined by the random mixture

$$F_\xi(s) = \sum_{k\geq 1} w_k F_{\xi_{(k)}}(s).$$

Here, weights $w_k = \frac{1}{\theta}\mathbb{E}\left(\xi_{(k)}\right)$ satisfy $w_k \in (0,1)$ and $\sum_{k\geq 1} w_k = 1$. With $\varphi_\xi(q) := \mathbb{E}\xi^q$, its moment function is equivalently given by

$$\varphi_\xi(q) = \frac{1}{\theta}\sum_{k\geq 1} \mathbb{E}\left(\xi_{(k)}\right) \varphi_{\xi_{(k)}}(q) \tag{3.21}$$

in terms of $\varphi_{\xi_{(k)}}(q) := \mathbb{E}\xi_{(k)}^q$, the moment functions of $\xi_{(k)}$.

Size-biased Picking from $\left(\xi_{(k)}, k \geq 1\right)$

Let η be a $(0,1)$ $-$valued random variable taking the value $\xi_{(k)}$ with probability $\frac{1}{\theta}\xi_{(k)}$ given $\left(\xi_{(k)}, k \geq 1\right)$. This random variable corresponds to a size-biased picking from $\left(\xi_{(k)}, k \geq 1\right)$. Its moment function is

$$\varphi_\eta(q) = \mathbb{E}\eta^q := \mathbb{E}\frac{1}{\theta}\sum_{k\geq 1} \xi_{(k)}\xi_{(k)}^q = \frac{1}{\theta}\phi(q+1) \tag{3.22}$$

for $q > q_* := \beta_* - 1 \in [-1, 0)$.

The waiting time paradox reads

$$\eta \succeq_{st} \xi, \tag{3.23}$$

a stochastic domination property translating the fact that in the size-biased picking procedure, large fragments are favored.

3.2 Spacings and Strong Partition of Unity: Normalizing

The partition $\left(\xi_{(k)};\ k \geq 1\right)$ of χ induces another natural partition of unity defined as follows. Define the incremental random variables $\widetilde{\xi}_k := \xi_{(k-1)} - \xi_{(k)}$ (with $\xi_{(0)} := 1$), $k \geq 1$. Then, $\left(\widetilde{\xi}_k, k \geq 1\right)$ defines a new sequence of $(0,1)$ −valued random variables with clearly $\sum_{k \geq 1} \widetilde{\xi}_k = 1$ (almost surely). The $\left(\widetilde{\xi}_k;\ k \geq 1\right)$ constitute a strong (almost sure) random partition of unity built on χ. Spacings between consecutive ordered energies sum up to 1, which is the top energy a single event can develop according to our assumptions. This model was first considered by [18] and reconsidered by [19] in the context of combinatorial structures.

That this construction is possible is indeed a consequence of Π being concentrated on $(0,1)$ leading to $(0,1)$ −valued $\xi_{(k)}$s. Note that there are no reasons, in general, for the $\widetilde{\xi}_k$s to be ordered either in the strict or weaker stochastic sense. The ordered version of $\left(\widetilde{\xi}_k;\ k \geq 1\right)$, say $\left(\widetilde{\xi}_{(k)};\ k \geq 1\right)$, is thus also of some interest.

This construction should not be confused with another partition of unity which can be defined from the system of ordered normalized random weights $\varsigma_{(k)} := \xi_{(k)}/\chi$, $k \geq 1$ satisfying $\sum_{k \geq 1} \varsigma_{(k)} \overset{d}{=} 1$ and $\varsigma_{(1)} > .. > \varsigma_{(k)} > ...$ For this kind of partition of unity, the condition that Lévy measure Π of χ be concentrated on $(0,1)$ is inessential. When Π is concentrated on $(0,\infty)$, one speaks of Poisson-Kingman partitions (see [20]). For instance, when $\Pi(dx) = \theta e^{-x}/x$, $x > 0$, is the Lévy measure for jumps of a Moran (gamma) subordinator $(\chi_t; t \geq 0)$, $\left(\varsigma_{(k)};\ k \geq 1\right)$ has Poisson-Dirichlet PD(θ) distribution, independent of $\chi = \chi_1$. For such problems, the joint law of $\left(\chi; \varsigma_{(k)},\ k = 1,..,l\right)$ for each $l \geq 1$ deserves some attention. They are given by Perman formulae (see [21], for additional details).

Strong Partition Function of Unity from χ: Structural Measure

Let

$$\widetilde{\phi}(\beta) := \mathbb{E}\sum_{k \geq 1} \widetilde{\xi}_k^{\beta}, \text{ with } \beta > \beta_* \in [0,1).$$

With $S_0 := 0$, averaging over $S_k \overset{d}{\sim} \text{gamma}(k)$, $k \geq 1$, $\widetilde{\phi}(\beta)$ can be obtained in general from

$$\widetilde{\phi}(\beta) = \sum_{k \geq 1} \int_0^{\infty} e^{-t}\mathbb{E}\left[\left(\overline{\Pi}^{-1}(S_{k-1}) - \overline{\Pi}^{-1}(S_{k-1}+t)\right)^{\beta}\right] dt, \tag{3.24}$$

recalling $S_k = S_{k-1} + T_k$ where S_{k-1} is independent of $T_k \overset{d}{\sim} \exp(1)$. The measure $\sigma(dx)$ such that $\widetilde{\phi}(\beta) = \int_0^1 x^{\beta}\sigma(dx)$ is called the structural measure of the partition $\left(\widetilde{\xi}_k, k \geq 1\right)$. With $\overline{\sigma}(x) := \int_x^1 \sigma(dz)$, recalling $\mathbb{P}(S_k \in ds) = \frac{1}{(k-1)!}s^{k-1}e^{-s}ds$, it can generally be obtained, after a change of variable, from

$$\overline{\sigma}(x) = \sum_{k \geq 1} \int_0^{\infty} e^{-t}\mathbb{P}\left[\overline{\Pi}^{-1}(S_{k-1}) - \overline{\Pi}^{-1}(S_{k-1}+t) > x\right] dt$$

$$= e^{-\overline{\Pi}(1-x)} + \sum_{k\geq 1} \mathbb{E}\left\{ e^{-\left[\overline{\Pi}\left(\overline{\Pi}^{-1}(S_k)-x\right)-S_k\right]};\ S_k < \overline{\Pi}(x) \right\}$$

$$= e^{-\overline{\Pi}(1-x)} - \sum_{k\geq 1} \frac{1}{k!} \int_x^1 e^{-\overline{\Pi}(z-x)} d\overline{\Pi}^k (z)$$

$$= e^{-\overline{\Pi}(1-x)} + \int_x^1 e^{-\left[\overline{\Pi}(z-x)-\overline{\Pi}(z)\right]} \pi(z)\, dz. \tag{3.25}$$

The random variable $\sum_{k\geq 1} \widetilde{\xi}_k^{\beta}$ is called the partition function of unity constructed from χ and σ defined by $\overline{\sigma}(x) = \sum_{k\geq 1} \mathbb{P}\left(\widetilde{\xi}_k > x\right)$ its structural measure.

Cutoff Considerations for Spacings Partition

1/ Overshoot. First, we note that, if $\overline{\xi}_k := 1 - \sum_{l=1}^{k} \widetilde{\xi}_l$ is the amount of space left vacant by the k first atoms of $\left(\widetilde{\xi}_k, k \geq 1\right)$, we get $\overline{\xi}_k = \xi_{(k)}$. This remark allows us to derive the following result.

Let $x \in (0,1)$ be some threshold value. Define

$$K(x) := \inf\left(k \geq 1 : \sum_{l=1}^{k} \widetilde{\xi}_l > x\right)$$

to be the first time the cumulated fragments size of $\left(\widetilde{\xi}_k, k \geq 1\right)$ exceeds x. Then,

$$K(x) \overset{d}{=} 1 + P_{\Lambda(x)}$$

where $P_{\Lambda(x)}$ is a Poisson distributed random variable with parameter $\Lambda(x) := \overline{\Pi}(1-x)$.

Indeed, $\mathbb{P}(K(x) > k) = \mathbb{P}\left(\sum_{l=1}^{k} \widetilde{\xi}_l \leq x\right) = \mathbb{P}\left(\xi_{(k)} > 1-x\right)$.

The random quantity $\sum_{l=1}^{K(x)} \widetilde{\xi}_l - x$ is the overshoot at x.

2/ Let $\mathcal{N}_+(\epsilon) := \sum_{k\geq 1} \mathbf{I}\left(\widetilde{\xi}_k > \epsilon\right)$ be the random number of spacings $\widetilde{\xi}_k$ exceeding $\epsilon \in (0,1)$. Then, from the above expression of $\overline{\sigma}(x)$ in Eq. (3.25), evaluated in a neighborhood of $x = 0$,

$$\mathbb{E}\mathcal{N}_+(\epsilon) = \overline{\sigma}(\epsilon) \sim_{\epsilon\downarrow 0} \overline{\Pi}(\epsilon),$$

and Chen-Stein methods for Poisson approximations of $\mathcal{N}_+(\epsilon)$ could be developed, in the spirit of [22].

However, $\mathbb{P}\left(\inf\left(l \geq 1 : \widetilde{\xi}_l \leq \epsilon\right) > k\right) = \mathbb{P}\left(\wedge_{l=1}^{k} \widetilde{\xi}_k > \epsilon\right)$ where $\wedge_{l=1}^{k} \widetilde{\xi}_l$ is the smallest term amongst $\left(\widetilde{\xi}_1, .., \widetilde{\xi}_k\right)$ and it is no longer true that $\mathcal{N}_+(\epsilon) = \inf\left(k \geq 1 : \widetilde{\xi}_k \leq \epsilon\right) - 1$ because the $\widetilde{\xi}_k$ are not ordered.

The contribution to total mass of those $\widetilde{\xi}_k$ above or below ϵ are respectively $1_+(\epsilon) := \sum_{k\geq 1} \widetilde{\xi}_k \mathbf{I}\left(\widetilde{\xi}_k > \epsilon\right)$ and $1_-(\epsilon) := \sum_{k\geq 1} \widetilde{\xi}_k \mathbf{I}\left(\widetilde{\xi}_k \leq \epsilon\right)$. The full laws of these quantities are difficult to obtain in general. Indeed,

$$\mathbb{E}\exp\left\{-\lambda\mathcal{N}_+(\epsilon)\right\} = \mathbb{E}\prod_{k\geq 1}\left(1 - \left(1 - e^{-\lambda}\right)\mathbf{I}\left(\widetilde{\xi}_k > \epsilon\right)\right)$$

$$\mathbb{E}\exp\left\{-\lambda 1_{\pm}(\epsilon)\right\} = \mathbb{E}\prod_{k\geq 1}\left(1-\left(1-e^{-\lambda\widetilde{\xi}_k}\right)\mathbf{I}\left(\widetilde{\xi}_k \gtreqless \epsilon\right)\right)$$

and the joint laws of the $\widetilde{\xi}_k$ are required. However, as $\epsilon \downarrow 0$, it still holds that

$$1_{+}(\epsilon) \xrightarrow{d} 1 \text{ and } \mathbb{E}1_{-}(\epsilon) \sim \epsilon^2 \sum_{k\geq 1} f_{\widetilde{\xi}_k}(\epsilon) = \epsilon^2\sigma(\epsilon) \sim \epsilon^2\pi(\epsilon).$$

Size-biased Picking from $\left(\widetilde{\xi}_k, k\geq 1\right)$

Let $\widetilde{\eta}$ be a $(0,1)-$valued random variable taking the value $\widetilde{\xi}_k$ with probability $\widetilde{\xi}_k$ given $\left(\widetilde{\xi}_k, k\geq 1\right)$. This random variable corresponds to a size-biased picking from $\left(\widetilde{\xi}_k, k\geq 1\right)$. Its moment function is

$$\varphi_{\widetilde{\eta}}(q) = \mathbb{E}\widetilde{\eta}^q := \mathbb{E}\sum_{k\geq 1}\widetilde{\xi}_k^{q+1} = \widetilde{\phi}(q+1) \tag{3.26}$$

for $q > q_* := \beta_* - 1 \in [-1, 0)$.

Just like for the pair η and ξ, the waiting time paradox reads $\widetilde{\eta} \succeq_{st} \widetilde{\xi}$.

4 Examples: Dickman partition and related ones

We start with a fundamental example in many respects, for which most computations can be painlessly achieved. We call it Dickman partition for reasons to appear later. The peculiarities of this model clearly appeared in the Examples developed to illustrate the general partition model under study in the previous Sections.

4.1 Dickman Partition

• Assume $\Pi(dx) = \frac{\theta}{x}\mathbf{I}_{x\in(0,1)}dx$. Then $\overline{\Pi}(x) = -\theta\log x$ and $\overline{\Pi}^{-1}(s) = e^{-s/\theta}$. The LST of χ in this case is

$$\begin{aligned}\mathbb{E}e^{-\lambda\chi} &= \exp\left\{-\theta\int_0^1 \frac{1-e^{-\lambda x}}{x}dx\right\}\\ &= \exp\left\{-\theta\int_0^\lambda \frac{1-e^{-\lambda'}}{\lambda'}d\lambda'\right\}.\end{aligned}$$

Let us now proceed with the detailed study of the multiplicative structure of this partitioning model.

Partition Function

In this example, we have $\xi_{(k)} = e^{-S_k/\theta} = \prod_{l=1}^k B_l$ where B_k are iid with beta$(\theta,1)$ law: $\mathbb{P}(B_1 \leq x) = x^\theta$. The $\xi_{(k)}$ are thus log-gamma(k,θ) distributed. We obtain

$$\mathbb{E}\xi_{(k)}^\beta = \left(\frac{\theta}{\theta+\beta}\right)^k \text{ and } \mathbb{E}\sum_{k\geq 1}\xi_{(k)}^\beta = \frac{\theta}{\beta} = \int_0^1 x^\beta\frac{\theta}{x}dx, \tag{4.1}$$

for $\beta > \beta_* = 0$. The structural measure is $\sigma(dx) = \frac{\theta}{x}dx$.

Note also that $-\frac{\theta}{k}\log\left(\xi_{(k)}\right) \to 1$ almost surely as $k \uparrow \infty$ so that $\xi_{(k)}$ goes to 0 exponentially fast with k.

Note that, exploiting the product decomposition of the $\xi_{(k)}$, if $\varsigma_{(1)} := \xi_{(1)}/\chi$ is the largest normalized fragment of the $\xi_{(k)}$'s, we get $\varsigma_{(1)} \overset{d}{=} 1/(1+\chi)$. Consequently,

$$\mathbb{E}e^{-\lambda/\varsigma_{(1)}} = \exp\left\{-\lambda - \theta\int_0^1 \frac{1-e^{-\lambda x}}{x}dx\right\} \tag{4.2}$$

is the infinitely divisible LST of $1/\varsigma_{(1)} > 1$.

Correlation Structure

The correlation structure of the Dickman partition has already been computed in a former example. The result is the expression of $C_l(q_1, q_2)$ in (3.5).

Typical and Size-biased Fragment Size

The size-biased picking random variable η from $\left(\xi_{(k)}; k \geq 1\right)$ has uniform law since

$$\mathbb{E}\eta^q := \mathbb{E}\frac{1}{\theta}\sum_{k\geq 1}\xi_{(k)}^{q+1} = 1/(q+1). \tag{4.3}$$

The typical fragment size ξ has law given by

$$\varphi_\xi(q) = \mathbb{E}\xi^q := \frac{1}{\theta}\sum_{k\geq 1}\mathbb{E}\xi_{(k)}\mathbb{E}\xi_{(k)}^q = \frac{\theta}{(1+\theta)q+\theta} \tag{4.4}$$

corresponding to a beta$\left(\frac{\theta}{1+\theta}, 1\right)$ distribution. It is true that $\eta \succeq_{st} \xi$ since $F_\eta(x) = x \leq F_\xi(x) = x^{\frac{\theta}{1+\theta}}$ for all $x \in [0,1]$.

Additional Properties: the Law of χ

With γ the Euler constant, the random variable χ has a density given by

$$f_\chi(x) = e^{-\gamma\theta}x^{\theta-1}\overline{F}_\theta(x)/\Gamma(\theta),\ x > 0 \tag{4.5}$$

where $\overline{F}_\theta(x) := \mathbb{P}(D > x)$ is the tail probability distribution function of Dickman random variable D

$$\overline{F}_\theta(x) = \mathbf{I}_{x\in[0,1)} + \mathbf{I}_{x\geq 1}\left(1 + \sum_{j=1}^{[x]}\frac{(-\theta)^j}{j!}\int_{\frac{1}{x}}^1 .. \int_{\frac{1}{x}}^1 \frac{\left(1-\sum_{l=1}^j z_l\right)_+^{\theta-1}}{\prod_{l=1}^j z_l}\prod_{l=1}^j dz_l\right)$$

with super-exponential von Mises tails $-\log\overline{F}_\theta(x) \sim_{x\uparrow\infty} x\log x$ (see [11]). When $x \geq 1$, the function $\overline{F}_\theta(x)$ is the solution to

$$x^\theta\overline{F}_\theta(x) = \theta\int_{x-1}^x z^{\theta-1}\overline{F}_\theta(z)\,dz.$$

The random variable $D>1$ turns out to be the reciprocal of the largest normalized jump of the Moran (gamma) subordinator ($D=1/\varsigma_{(1)}$).

This relationship between $f_{\chi}(x)$ and $\overline{F}_{\theta}(x)$ may be seen to be a direct consequence of the identity

$$\exp\left(-\theta\left(\int_{1}^{\infty}\frac{e^{-\lambda x}}{x}dx+\log\lambda+\gamma\right)\right)=\exp\left(-\theta\int_{0}^{1}\frac{1-e^{-\lambda x}}{x}dx\right)$$

involving the exponential integral function $\mathrm{Ei}(\lambda)=\int_{1}^{\infty}\frac{e^{-\lambda x}}{x}dx$ (see [11]). For these connections with Dickman's function, we shall say from (4.5) that χ follows Dickman distribution.

Moment Function of χ

The moment function of χ, say $\mathbb{E}\chi^{q}=\int_{0}^{\infty}x^{q}f_{\chi}(x)\,dx$, is defined for $q>-\theta$, with

$$\mathbb{E}\chi^{q}=\frac{e^{-\gamma\theta}}{\Gamma(\theta)(q+\theta)}\mathbb{E}D^{q+\theta}$$

It can be computed as follows: first, $\chi\overset{d}{=}\xi_{(1)}\left(1+\chi^{'}\right)$ where $\chi^{'}\overset{d}{=}\chi$ is independent of $\xi_{(1)}\overset{d}{\sim}\text{beta}(\theta,1)$ with $\mathbb{E}\xi_{(1)}^{q}=\theta/(\theta+q)$. Thus, χ is a Vervaat perpetuity of a special type.

Let $\mu_{n}(\theta):=\mathbb{E}\xi_{(1)}^{n}$; from this, using the binomial identity, the integral moments $m_{n}(\theta)$ of χ are first obtained recursively by $m_{0}(\theta)=1$ and

$$m_{n}(\theta)=\frac{\mu_{n}(\theta)}{1-\mu_{n}(\theta)}\sum_{p=0}^{n-1}\binom{n}{p}m_{p}(\theta),\ n\geq1, \tag{4.6}$$

with $\frac{\mu_{n}(\theta)}{1-\mu_{n}(\theta)}=\frac{\theta}{n}$. Hence

$$\begin{aligned}
m_{1}(\theta)&=\theta,\\
m_{2}(\theta)&=\frac{\theta}{2}(1+2m_{1}(\theta))=\frac{\theta}{2}+\theta^{2},\\
m_{3}(\theta)&=\frac{\theta}{3}(1+3m_{1}(\theta)+3m_{2}(\theta))=\frac{\theta}{3}+\frac{3\theta^{2}}{2}+\theta^{3}
\end{aligned}$$

are the three first nested moments. Searching for solutions under the polynomial form

$$m_{n}(\theta)=\sum_{k=1}^{n}b_{k,n}\theta^{k},\ n\geq1$$

we can identify the coefficients $b_{k,n}$, $k=2,..,n$ as the ones solving ($b_{1,n}=1/n$) the Bell numbers-like triangular recurrence

$$b_{k,n}=\frac{1}{n}\sum_{p=k-1}^{n-1}\binom{n}{p}b_{k-1,p},\ k=2,..,n.$$

Next, with $(q)_n := q\,(q-1)\,..\,(q-n+1)$, the full expression of $\mathbb{E}\chi^q$ is

$$\mathbb{E}\chi^q = \frac{\theta}{\theta+q}\left(1+\sum_{n\geq 1}\frac{(q)_n}{n!}m_n(\theta)\right),\ q>-\theta, \tag{4.7}$$

translating the identity $\mathbb{E}\chi^q = \frac{\theta}{\theta+q}\mathbb{E}(1+\chi)^q$. Incidentally,

$$\mathbb{E}D^q = \Gamma(\theta+1)\,e^{\gamma\theta}\left(1+\sum_{n\geq 1}\frac{(q-\theta)_n}{n!}m_n(\theta)\right), q>0$$

is the moment function of $D = 1/\varsigma_{(1)}$.

Let $\chi_{1/\beta} := \sum_{k\geq 1}\xi_{(k)}^{\beta}$, with $\beta>0$. From Campbell formula,

$$\mathbb{E}e^{-\lambda\chi_{1/\beta}} = \exp\left\{-\frac{1}{\beta}\int_0^1\left(1-e^{-\lambda x}\right)\Pi(dx)\right\} = \left(\mathbb{E}e^{-\lambda\chi}\right)^{1/\beta}. \tag{4.8}$$

The moment function $\mathbb{E}\chi_{1/\beta}^q$ can thus readily be obtained from the one of $\mathbb{E}\chi^q$ while operating the substitution $\theta\to\theta/\beta$ in the obtained formula. Hence

$$\mathbb{E}\chi_{1/\beta}^q = \frac{\theta}{\theta+\beta q}\left(1+\sum_{n\geq 1}\frac{(q)_n}{n!}m_n(\theta/\beta)\right),\ q>-\theta/\beta \tag{4.9}$$

is the moment function of the partition function $\sum_{k\geq 1}\xi_{(k)}^{\beta}$. Note that $\mathbb{E}\chi_{1/\beta} = \frac{\theta}{\theta+\beta}(1+\theta/\beta) = \theta/\beta$, as required. This constitutes a complementary information to the one encoded in the LST of $\sum_{k\geq 1}\xi_{(k)}^{\beta}$. This suggests the following additional construction.

Rényi's Weighted Averages

Let $\varsigma_{(k)} := \xi_{(k)}/\chi,\ k\geq 1$ be a system of normalized random weights. With $\beta>-1$, define the random Rényi $\beta-$average $[\xi]_\beta$ (with random weights $\varsigma_{(k)}$) of the $\left(\xi_{(k)}, k\geq 1\right)$ to be $(\beta>-1)$

$$[\xi]_\beta := \left(\sum_{k\geq 1}\varsigma_{(k)}\xi_{(k)}^{\beta}\right)^{1/\beta} = \left(\frac{1}{\chi}\sum_{k\geq 1}\xi_{(k)}^{\beta+1}\right)^{1/\beta} = \left(\frac{\chi_{1/(\beta+1)}}{\chi}\right)^{1/\beta}.$$

This random variable is $(0,1)-$valued when $\beta>-1$ and when $\beta\leq -1$, it degenerates to 0. The $2-$average $\langle\xi\rangle_2$ is often considered, but $[\xi]_0 := \lim_{\beta\uparrow 0}[\xi]_\beta = \prod_{k\geq 1}\xi_{(k)}^{\varsigma_k}$ is also sometimes of interest. The function $\beta\to[\xi]_\beta$ is non-decreasing with β and $[\xi]_{\beta_1} > [\xi]_{\beta_2}$ if $-1<\beta_2<\beta_1$.

Note that $\mathbb{E}[\xi]_\beta^q$ is also $\mathbb{E}\left[e^{-qH_\beta}\right]$ where $H_\beta = -\log[\xi]_\beta$ is the random Rényi $\beta-$entropy of the sequence $\left(\xi_{(1)},..,\xi_{(k)},..\right)$, with, in particular, $H_0 = -\log[\xi]_0 = -\sum_{k\geq 1}\varsigma_k\log\xi_{(k)}$, related to Shannon entropy.

The computation of its moment function turns out to be difficult as it stands. Indeed, recalling that $(\chi_t, t\geq 0)$ is a process with stationary independent increments,

the joint moment function of $\chi_{1/(\beta+1)}$ and $\chi = \chi_1$ would first be required.

We shall rather consider the simpler Rényi $\beta-$average (with deterministic weights w_k) of the $\left(\xi_{(k)}, k \geq 1\right)$ to be

$$\langle \xi \rangle_\beta := \left(\sum_{k \geq 1} w_k \xi_{(k)}^\beta \right)^{1/\beta} \quad \text{where } w_k = \frac{\mathbb{E}\xi_{(k)}}{\theta} \text{ and } \beta > 0. \tag{4.10}$$

It may be checked that the range of the random variable $\langle \xi \rangle_\beta$ is $[0,1]$ when $\beta \in (0, \infty)$ which we shall limit ourselves to. Define the weighted sum

$$W := \sum_{k \geq 1} w_k \xi_{(k)}^\beta$$

Recalling $w_k = \frac{1}{\theta}\left(\frac{\theta}{\theta+1}\right)^k$ and $\xi_{(k)} = \prod_{l=1}^k B_k$ where B_k are iid with $\xi_{(1)} = B_1 \overset{d}{\sim}$ beta$(\theta, 1)$ law, we have

$$W \overset{d}{=} \frac{\xi_{(1)}^\beta}{\theta+1}\left(1 + \theta W'\right)$$

where $W' \overset{d}{=} W$ is independent of $\xi_{(1)}^\beta \overset{d}{\sim}$ beta$\left(\frac{\theta}{\beta}, 1\right)$. Let $\mu_n(\theta, \beta) := \frac{1}{(\theta+1)^n}\mathbb{E}\xi_{(1)}^{n\beta} = \frac{1}{(\theta+1)^n}\frac{\theta}{\theta+n\beta}$; from this, using the binomial identity, the integral moments $m_n(\theta, \beta)$ of W are first obtained recursively by $m_0(\theta, \beta) = 1$ and

$$m_n(\theta, \beta) = \frac{\mu_n(\theta, \beta)}{1 - \mu_n(\theta, \beta)} \sum_{p=0}^{n-1} \binom{n}{p} \theta^p m_p(\theta, \beta), \; n \geq 1,$$

with $\frac{\mu_n(\theta)}{1-\mu_n(\theta)} = \frac{\theta}{(\theta+1)^n(\theta+n\beta)-\theta}$. From this

$$\mathbb{E}W^q = \theta \frac{1 + \sum_{n \geq 1} \frac{(q)_n}{n!} \theta^n m_n(\theta, \beta)}{(1+\theta)^q (\theta + \beta q)} \tag{4.11}$$

and

$$\mathbb{E}\langle \xi \rangle_\beta^q = \mathbb{E}W^{q/\beta} = \theta \frac{1 + \sum_{n \geq 1} \frac{(q/\beta)_n}{n!} \theta^n m_n(\theta, \beta)}{(1+\theta)^{q/\beta}(\theta + q)}. \tag{4.12}$$

4.2 Spacings: an Alternative Construction of Poisson-Dirichlet Partition

Defining spacings to be $\widetilde{\xi}_k := \xi_{(k-1)} - \xi_{(k)}$ (with $\xi_{(0)} := 1$), we clearly have

$$\widetilde{\xi}_k \overset{d}{=} \prod_{l=1}^{k-1} \overline{v}_l v_k, \; k \geq 1 \tag{4.13}$$

where v_k are iid with beta$(1, \theta)$ law: $\mathbb{P}(v_1 \leq x) = (1-x)^\theta$. Thus $\left(\widetilde{\xi}_k, k \geq 1\right)$ has GEM(θ) distribution with $\sum_{k \geq 1} \widetilde{\xi}_k = 1$.

In this particular example, one can also check that $\widetilde{\xi}_1 \succeq_{st} .. \succeq_{st} \widetilde{\xi}_k \succeq_{st} ..$ and the $\left(\widetilde{\xi}_k, k \geq 1\right)$ are arranged in stochastic descending order. Finally, as is well-known, the ordered version $\left(\widetilde{\xi}_{(k)}, k \geq 1\right)$ of $\left(\widetilde{\xi}_k, k \geq 1\right)$ has Poisson-Dirichlet PD(θ) $-$distribution and $\left(\widetilde{\xi}_k, k \geq 1\right)$, as a size-biased permutation from PD(θ), is invariant under size-biased permutation.

Correlation Structure

With $w_k := \mathbb{E}\widetilde{\xi}_k$, the second order quantity to consider here is

$$\widetilde{C}_l\left(q_1, q_2\right) := \mathbb{E}\left(\sum_{k\geq 1} w_k \widetilde{\xi}_k^{q_1} \widetilde{\xi}_{k+l}^{q_2}\right). \tag{4.14}$$

From the multiplicative RAM structure of the GEM(θ) partition,

$$\begin{aligned}
\widetilde{C}_l\left(q_1, q_2\right) &= \frac{1}{1+\theta}\sum_{k\geq 1}\left(\frac{\theta}{1+\theta}\right)^{k-1}\mathbb{E}\left(\prod_{l=1}^{k-1}\overline{v}_l^{q_1+q_2} v_k^{q_1}\overline{v}_k^{q_2}\prod_{l=k+1}^{k+l-1}\overline{v}_l^{q_2} v_{k+l}^{q_2}\right) \\
&= \sum_{k\geq 1}\left(\frac{\theta}{1+\theta}\right)^{k}\left(\frac{\theta}{\theta+q_1+q_2}\right)^{k-1}\frac{\Gamma\left(1+q_1\right)\Gamma\left(\theta+q_2\right)}{\Gamma\left(1+\theta+q_1+q_2\right)}\times \\
&\qquad \left(\frac{\theta}{\theta+q_2}\right)^{l-1}\left(\frac{\Gamma\left(1+q_2\right)\Gamma\left(1+\theta\right)}{\Gamma\left(1+\theta+q_2\right)}\right) \\
&= K\left(q_1, q_2\right)\left(\frac{\theta}{\theta+q_2}\right)^{l} \qquad (4.15)
\end{aligned}$$

where

$$K\left(q_1, q_2\right) = \frac{\Gamma\left(1+q_1\right)\Gamma\left(1+q_2\right)\Gamma\left(1+\theta\right)}{\Gamma\left(\theta+q_1+q_2\right)\left[\theta+\left(\theta+1\right)\left(q_1+q_2\right)\right]} \tag{4.16}$$

is defined for $\{q_1+q_2 > -\theta/\left(\theta+1\right); \left(q_1, q_2\right) > -1\}$. The definition domain of $\widetilde{C}_l\left(q_1, q_2\right)$ therefore is $\{q_1+q_2 > -\theta/\left(\theta+1\right); q_1 > -1; q_2 > -\min\left(1, \theta\right)\}$. This should be compared with the expression of $C_l\left(q_1, q_2\right)$ in (3.5), dealing with the unnormalized case.

Cutoff Considerations for GEM(θ) Partition

Let $\mathcal{N}_+\left(\epsilon\right) := \sum_{k\geq 1}\mathbf{I}\left(\widetilde{\xi}_k > \epsilon\right)$ be the random number of spacings $\widetilde{\xi}_k$ exceeding $\epsilon \in (0, 1)$. Then, using the RAM structure of $\left(\widetilde{\xi}_k; k \geq 1\right)$, we have

$$\mathcal{N}_+\left(\epsilon\right) \overset{d}{=} \mathbf{I}\left(v_1 > \epsilon\right) + \mathcal{N}_+^{'}\left(\epsilon/\overline{v}_1\right)\mathbf{I}_{\overline{v}_1 > \epsilon}$$

where $\mathcal{N}_+^{'}\left(.\right)$ is a statistical copy of $\mathcal{N}_+\left(.\right)$. Here $v_1 \overset{d}{\sim}$ beta$(1, \theta)$ and $\overline{v}_1 := 1 - v_1$ is independent of $\mathcal{N}_+^{'}\left(.\right)$. In particular, if $n_+^{(p)}\left(\epsilon\right) = \mathbb{E}\mathcal{N}_+\left(\epsilon\right)$ is the order-p moment of

$\mathcal{N}_+(\epsilon)$, we have the following recurrence

$$n_+^{(1)}(\epsilon) = (1-\epsilon)^\theta + \theta \int_\epsilon^1 v^{\theta-1} n_+^{(1)}(\epsilon/v)\, dv \text{ and, with } p \geq 2:$$

$$n_+^{(p)}(\epsilon) = (1-\epsilon)^\theta + \theta \sum_{k=1}^{p-1} \binom{p}{k} \mathbf{I}_{\epsilon<1/2} \int_\epsilon^{1-\epsilon} v^{\theta-1} n_+^{(k)}(\epsilon/v)\, dv$$
$$+\theta \int_\epsilon^1 v^{\theta-1} n_+^{(p)}(\epsilon/v)\, dv.$$

Let

$$\widehat{n}_+^{(1)}(q) := \int_0^1 \epsilon^q n_+^{(1)}(\epsilon)\, d\epsilon.$$

From the first recurrence on $n_+^{(1)}(\epsilon)$, recalling

$$\int_0^1 \epsilon^q (1-\epsilon)^\theta\, d\epsilon = \frac{\Gamma(q+1)\Gamma(\theta+1)}{\Gamma(q+\theta+2)}$$

and using an integration by part, we obtain explicitly

$$\widehat{n}_+^{(1)}(q) = \frac{\Gamma(q+1)\Gamma(\theta+1)}{\Gamma(q+\theta+1)(q+1)}, \tag{4.17}$$

leading to $n_+^{(1)}(\epsilon) = \theta \int_\epsilon^1 z^{-1}(1-z)^{\theta-1}\, dz$. The function $\widehat{n}_+^{(1)}(q)$ has a double pole at $q=-1$ and $\widehat{n}_+^{(1)}(q) \sim_{q=-1} \theta(q+1)^{-2}$, showing, as required, from singularity analysis, that $n_+^{(1)}(\epsilon) \sim_{\epsilon\downarrow 0} -\theta \log \epsilon$.

Similarly, considering the contribution to total mass of those $\widetilde{\xi}_k$ above or below ϵ, which are respectively $1_+(\epsilon) := \sum_{k\geq 1} \widetilde{\xi}_k \mathbf{I}\left(\widetilde{\xi}_k > \epsilon\right)$ and $1_-(\epsilon) := \sum_{k\geq 1} \widetilde{\xi}_k \mathbf{I}\left(\widetilde{\xi}_k \leq \epsilon\right)$, it holds that

$$1_+(\epsilon) \overset{d}{=} v_1 \mathbf{I}(v_1 > \epsilon) + \overline{v}_1 1_+^{'}(\epsilon/\overline{v}_1)\, \mathbf{I}_{\overline{v}_1 > \epsilon}$$
$$1_-(\epsilon) \overset{d}{=} v_1 \mathbf{I}(v_1 \leq \epsilon) + \overline{v}_1 1_-^{'}(\epsilon/\overline{v}_1)\, \mathbf{I}_{\overline{v}_1 > \epsilon}$$

where $1_\pm^{'}(.)$ are statistical copies of $1_\pm(.)$ respectively, independent of $\overline{v}_1$. Let $x_\pm(\epsilon) := \mathbb{E}[1_\pm(\epsilon)]$ stand for the mean values. Then, with

$$a_+(\epsilon) := \theta \int_\epsilon^1 v(1-v)^{\theta-1}\, dv \text{ and } a_-(\epsilon) := \theta \int_0^\epsilon v(1-v)^{\theta-1}\, dv,$$

we have

$$x_\pm(\epsilon) = a_\pm(\epsilon) + \theta \int_\epsilon^1 v^\theta x_\pm(\epsilon/v)\, dv.$$

Defining $\widehat{x}_\pm(q) := \int_0^1 \epsilon^q x_\pm(\epsilon)\, d\epsilon$, similar computations show that
1/

$$\widehat{x}_+(q) = \frac{\Gamma(q+1)\Gamma(\theta+1)}{\Gamma(q+\theta+2)} \tag{4.18}$$

and $\widehat{x}_+(q)$ has a simple dominant pole at $q=-1$ with $\widehat{x}_+(q) \sim_{q=-1} (q+1)^{-1}$, showing, as required, from singularity analysis, that $x_+(\epsilon) \sim_{\epsilon\downarrow 0} 1$. More precisely, inverting (4.18), $x_+(\epsilon) = (1-\epsilon)^\theta$ and $x_+(\epsilon) \sim_{\epsilon\downarrow 0} 1-\theta\epsilon$.

2/

$$\widehat{x}_{-}(q)=\frac{(q+\theta+2)\,\Gamma(\theta+1)}{(q+1)(q+2)}\left\{\frac{\Gamma(q+\theta+3)-\Gamma(\theta+2)\,\Gamma(q+3)}{\Gamma(\theta+2)\,\Gamma(q+\theta+3)}\right\} \tag{4.19}$$

and $\widehat{x}_{-}(q)$ has a simple dominant pole at $q=-2$ with $\widehat{x}_{-}(q)\sim_{q=-2}\frac{\theta^2}{\theta+1}(q+2)^{-1}$, showing, as required, from singularity analysis, that $x_{-}(\epsilon)\sim_{\epsilon\downarrow 0}\frac{\theta^2}{\theta+1}\epsilon$.

Note that the average values $x_{\pm}(\epsilon)$ are known explicitly through inverting $\widehat{x}_{\pm}(q)$, even if ϵ is not small (in particular $x_{+}(\epsilon)=(1-\epsilon)^{\theta}$ and $x_{-}(\epsilon)=1-\frac{\theta}{\theta+1}\epsilon-(1-\epsilon)^{\theta}$). Singularity analysis of $\widehat{x}_{\pm}(q)$ gives the expected small$-\epsilon$ behavior of $x_{\pm}(\epsilon)$.

Similar recurrence on higher-order moments of $1_{\pm}(\epsilon)$ could be obtained.

Weighted partition: modulation

Assume $\left(\widetilde{\xi}_{(k)};k\geq 1\right)\overset{d}{\sim}\text{GEM}(\theta)$. Let $(\mu_k;k\geq 1)$ be a sequence of iid non-negative random variables. We shall investigate some properties of the weighted or modulated random sequence $\left(\mu_k\widetilde{\xi}_{(k)};k\geq 1\right)$ as a new random transformed partition of $\chi_{\mu_1}:=\sum_{k\geq 1}\mu_k\widetilde{\xi}_{(k)}$. Plainly, we have

$$\chi_{\mu_1}\overset{d}{=}\mu_1 v_1+\overline{v}_1\chi'_{\mu_1}$$

where $\chi'_{\mu_1}\overset{d}{=}\chi_{\mu_1}$ is independent of $\overline{v}_1\overset{d}{\sim}\text{beta}(\theta,1)$. Assume $c_k:=\mathbb{E}\left[\mu_1^k\right]<\infty$ for all $k\geq 0$. Then, with $m_n:=\mathbb{E}\left[\chi_{\mu_1}^n\right]$, we have

$$m_n=\sum_{k=0}^{n-1}\binom{n}{k}c_{n-k}\mathbb{E}\left[v_1^{n-k}\overline{v}_1^k\right]m_k+\mathbb{E}\left[\overline{v}_1^n\right]m_n.$$

Recalling $\mathbb{E}\left[v_1^{n-k}\overline{v}_1^k\right]=\theta\left[(n-k)!\Gamma(\theta+k)\right]/\Gamma(\theta+n+1)$ and $\mathbb{E}\left[\overline{v}_1^n\right]=\theta/(\theta+n)$, setting $h_n:=m_n\Gamma(\theta+n)/n!$, we obtain the convolution-like recurrence $(n\geq 1)$

$$nh_n=\theta\sum_{k=0}^{n-1}c_{n-k}h_k.$$

If

$$h(u):=\sum_{n\geq 0}h_n u^n$$

is the generating function of $(h_n;n\geq 1)$, with $c(u):=\sum_{n\geq 0}c_n u^n$, we obtain $uh'(u)+\theta h(u)=\theta c(u)\,h(u)$, leading explicitly to

$$h(u)=\Gamma(\theta)\exp\left\{-\theta\int_0^u\frac{1-c(v)}{v}dv\right\}. \tag{4.20}$$

In more details, with $\widetilde{c}_k:=(k-1)!c_k$, with $n\geq 1$, we finally get

$$m_n=\frac{\Gamma(\theta)\,n!}{\Gamma(\theta+n)}\sum_{k=1}^{n}(-\theta)^k B_{n,k}\left(\widetilde{c}_1,\widetilde{c}_2,...,\widetilde{c}_{n-k+1}\right)$$

in terms of Bell polynomials.

The moments of the partition function

$$\chi_{\mu_1}(\beta) := \sum_{k\geq 1}\left[\mu_k \widetilde{\xi}_{(k)}\right]^{\beta}$$

of $\left(\mu_k \widetilde{\xi}_{(k)}; k \geq 1\right)$ can be obtained similarly, observing that, with $\overline{v}_1^{\beta} \overset{d}{\sim} \text{beta}(\theta/\beta, 1)$

$$\chi_{\mu_1}(\beta) \overset{d}{=} \mu_1^{\beta} v_1^{\beta} + \overline{v}_1^{\beta}\chi_{\mu_1}(\beta).$$

Example:

When $(\mu_k; k \geq 1)$ is a sequence of iid non-negative random variables drawn for Bernoulli(p) distribution, with $c_k = p$, $k \geq 1$, one can check that

$$m_n = \frac{\Gamma(p\theta + n)\Gamma(\theta)}{\Gamma(p\theta)\Gamma(\theta + n)},\ n \geq 1$$

solves the recurrence. This shows that $\chi_{\mu_1} := \sum_{k\geq 1}\mu_k \widetilde{\xi}_{(k)}$ has beta$(p\theta, (1-p)\theta)$ distribution in this case (see [23]). $\square$

Structural Measure

With $\beta > \beta_* = 0$, we also have in this case

$$\widetilde{\phi}(\beta) = \frac{\Gamma(\beta)\Gamma(\theta + 1)}{\Gamma(\theta + \beta)} = \int_0^1 x^{\beta}\frac{\theta}{x}(1-x)^{\theta - 1}dx. \tag{4.21}$$

The measure $\sigma(dx) := \frac{\theta}{x}(1-x)^{\theta-1}\mathbf{I}_{x\in(0,1)}dx$ is the structural measure of the GEM(θ) partition.

Typical Fragment Size $\widetilde{\xi}$ from $\left(\widetilde{\xi}_k, k \geq 1\right)$

Its moment function is given by

$$\varphi_{\widetilde{\xi}}(q) = \sum_{k\geq 1}\mathbb{E}\left(\widetilde{\xi}_k\right)\varphi_{\widetilde{\xi}_k}(q)$$

where $\varphi_{\widetilde{\xi}_k}(q) := \mathbb{E}\widetilde{\xi}_k^q = \frac{\Gamma(q+1)\Gamma(\theta+1)}{\Gamma(\theta+q+1)}\left(\frac{\theta}{\theta+q}\right)^{k-1}$ is the moment function of $\widetilde{\xi}_k$. As a result

$$\varphi_{\widetilde{\xi}}(q) = \frac{\Gamma(q+1)\Gamma(\theta+1)}{\Gamma(\theta+q+1)(\theta+1)}\sum_{k\geq 1}\left(\frac{\theta^2}{(\theta+1)(\theta+q)}\right)^{k-1}$$

$$= \frac{\Gamma(q+1)\Gamma(\theta+1)}{\Gamma(\theta+q+1)}\frac{\theta+q}{\theta+(\theta+1)q} \tag{4.22}$$

showing that $\widetilde{\xi} \stackrel{d}{=} \widetilde{\eta} \cdot R$ where $\widetilde{\eta} \stackrel{d}{\sim} \text{beta}(1,\theta)$ is independent of the $[0,1]$ −valued random variable R, with moment function $\mathbb{E}R^q = \frac{\theta+q}{\theta+(\theta+1)q}$. This interprets as follows: let B be a Bernoulli random variable with parameter $\frac{1}{\theta+1}$ and $C \stackrel{d}{\sim} \text{beta}\left(\frac{\theta}{\theta+1}, 1\right)$ a random variable on $[0,1]$, independent of B. Then, R is a $[0,1]$-valued random variable satisfying

$$R \stackrel{d}{=} B + (1-B) \cdot C$$

Indeed,

$$\mathbb{E}R^q = \frac{1}{\theta+1} + \frac{\theta}{\theta+1}\frac{\theta/(\theta+1)}{\theta/(\theta+1)+q} = \frac{\theta+q}{\theta+(\theta+1)q}.$$

Size-biased Picking $\widetilde{\eta}$ from $\left(\widetilde{\xi}_k, k \geq 1\right)$

Its moment function is

$$\mathbb{E}\widetilde{\eta}^q := \mathbb{E}\sum_{k\geq 1} \widetilde{\xi}_k^{q+1} = \widetilde{\phi}(q+1) = \frac{\Gamma(q+1)\Gamma(\theta+1)}{\Gamma(\theta+q+1)} \tag{4.23}$$

which is the moment function of a beta$(1,\theta)$ distributed random variable (as required from the size-biased picking invariance property of $\left(\widetilde{\xi}_k, k \geq 1\right)$).

The waiting time paradox

$$\widetilde{\eta} \succeq_{st} \widetilde{\xi}$$

is clearly satisfied from the decomposition $\widetilde{\xi} \stackrel{d}{=} \widetilde{\eta} \cdot R$.

In the next two examples, computations on spacings are quite involved. We skip them, focusing on the simplest aspects.

4.3 Two Additional Examples

We briefly sketch some properties of related partition models.

• Let $\alpha \in (0,1)$ and put $\overline{\alpha} := 1-\alpha$. Assume $\Pi(dx) = \theta\overline{\alpha}x^{-(1+\alpha)}\mathbf{I}_{x\in(0,1)}dx$. Then, with $a = \frac{\theta\overline{\alpha}}{\alpha} > 0$,

$$\overline{\Pi}(x) = a\left(x^{-\alpha}-1\right) \text{ and } \overline{\Pi}^{-1}(s) = (1+s/a)^{-1/\alpha}.$$

We have $\xi_{(k)} = (1+S_k/a)^{-1/\alpha}$ and

$$\mathbb{E}\sum_{k\geq 1}\xi_{(k)}^{\beta} = \int_0^1 x^{\beta}\frac{\theta\overline{\alpha}}{x^{1+\alpha}}dx = \frac{\theta\overline{\alpha}}{\beta-\alpha},$$

for $\beta > \alpha = \beta_* > 0$. Note that $\frac{a}{k}\left(\xi_{(k)}^{-\alpha}-1\right) \to 1$ almost surely as $k \uparrow \infty$ and $\xi_{(k)} \sim \left(\frac{k}{a}\right)^{-1/\alpha}$ goes to 0 algebraically fast with k (like $k^{-1/\alpha}$) in this case. Spacings are more involved.

- Let $\alpha \in (0,1)$, put $\overline{\alpha} := 1-\alpha$ and assume

$$\Pi(dx) = \frac{\theta}{\Gamma(\alpha)\Gamma(\overline{\alpha})} x^{-(1+\alpha)} (1-x)^{\alpha-1} \mathbf{I}_{x\in(0,1)} dx$$

with $\int_0^1 x\Pi(dx) = \theta$. We have

$$\mathbb{E}\sum_{k\geq 1} \xi_{(k)}^{\beta} = \frac{\theta}{\Gamma(\alpha)\Gamma(\overline{\alpha})} \int_0^1 x^{(\beta-\alpha)-1} (1-x)^{\alpha-1} dx = \frac{\theta\Gamma(\beta-\alpha)}{\Gamma(\overline{\alpha})\Gamma(\beta)},$$

for $\beta > \alpha = \beta_* > 0$.

When $\alpha = 1/2$, $\overline{\Pi}(x) = \frac{2\theta}{\pi}\left(\frac{1-x}{x}\right)^{1/2}$ and $\overline{\Pi}^{-1}(s) = \left(1+\left(\frac{\pi s}{2\theta}\right)^2\right)^{-1}$. We have $\xi_{(k)} = \left(1+\left(\frac{\pi S_k}{2\theta}\right)^2\right)^{-1}$ and $\frac{2\theta}{\pi k}\left(\xi_{(k)}^{-1}-1\right)^{1/2} \to 1$ almost surely as $k \uparrow \infty$. As a result, $\xi_{(k)} \sim \left(\frac{2\theta}{\pi k}\right)^2$ goes slowly to 0 (like k^{-2}) in this case. The associated size-biased picking random variable η follows the Arcsine(1/2) law.

In both examples, an α larger than 1 would violate the condition that Π is a Lévy measure.

5 The Bounded Partition Model with a Poissonian Number of Fragments

We finally briefly show that the partitioning model based on Lévy measure concentrated on $(0,1)$ is also of some statistical relevance in the case of a bounded Lévy measure for jumps. This model does not seem to have received attention in the literature.

In the bounded case, let $\mu := \overline{\Pi}(0) < \infty$. In this case

$$\mathbb{E}e^{-\lambda\chi} = \exp\left\{-\mu\left(1-\int_0^1 e^{-\lambda x} F(dx)\right)\right\} \tag{5.1}$$

where $F(dx) = \Pi(dx)/\mu$ is a probability distribution with mean value $\theta/\mu < 1$.

5.1 Poisson Partition of χ: the Model

Hence, with u_k, $k \geq 1$ an iid $(0,1)-$valued uniform sequence, P_μ a Poisson random variable with intensity μ, $\chi = \sum_{k=1}^{P_\mu} \overline{F}^{-1}(u_k)$ belongs to the class of compound Poisson random variables. Stated differently

$$\chi \overset{d}{=} \sum_{k=1}^{P_\mu} \overline{F}^{-1}\left(u_{(k),P_\mu}\right) \tag{5.2}$$

where $u_{(1),P_\mu} < .. < u_{(P_\mu),P_\mu}$ is obtained from sample $u_1,..,u_{P_\mu}$ while ordering the constitutive terms. We note that χ has an atom at $\chi = 0$ with probability $e^{-\mu}$ and that there are finitely many (Poissonian) fragments. This partition is also $\chi \overset{d}{=}$

$\sum_{k=1}^{P_\mu} \overline{\Pi}^{-1}\left(\mu u_{(k),P_\mu}\right)$ where, when $\mu \uparrow \infty$, $P_\mu \overset{a.s.}{\to} \infty$ and $\left(\mu u_{(1),P_\mu}, .., \mu u_{(k),P_\mu}\right) \overset{d}{\to} (S_1, .., S_k, ..)$ a Poisson point process on $\mathbb{R}^+$.

Defining $\xi_{(k),P_\mu} := \overline{F}^{-1}\left(u_{(k),P_\mu}\right)$, with $\xi_{(1),P_\mu} > .. > \xi_{(P_\mu),P_\mu}$, we get similarly

$$\phi(\beta) = \mathbb{E}\sum_{k=1}^{P_\mu} \xi_{(k),P_\mu}^{\beta} = \int_0^1 x^\beta \Pi(dx) \tag{5.3}$$

and the structural measure is $\Pi(dx) = \mu F(dx)$ which is bounded with $\int_0^1 x\Pi(dx) = \theta$.

Remark: Assume that the distribution F is the one of a beta$(\theta/(\mu-\theta), 1)$, $\mu > \theta$, with mean value θ/μ. Then

$$\overline{\Pi}(x) := \mu\overline{F}(x) = \mu\left(1 - x^{\theta/(\mu-\theta)}\right) \to_{\mu\uparrow\infty} -\theta\log x$$

which is the Lévy-measure tail of the limiting Dickman model discussed in Section 4.1. $\square$

Typical Fragment Size from $\left(\xi_{(k),P_\mu}, k = 1, .., P_\mu\right)$

In this case, the typical fragment size from $\left(\xi_{(k),P_\mu}, k = 1, .., P_\mu\right)$ clearly is the $(0,1)$ −valued random variable, say ξ, with probability distribution $F_\xi(x) = F(x)$ with $\mathbb{E}\xi = \theta/\mu$.

Size-biased Picking from $\left(\xi_{(k),P_\mu}, k = 1, .., P_\mu\right)$

Let η be a $(0,1)$ −valued random variable taking the value $\xi_{(k),P_\mu}$ with probability $\frac{1}{\theta}\xi_{(k),P_\mu}$ given $\left(\xi_{(k),P_\mu}, k = 1, .., P_\mu\right)$. This random variable corresponds to a size-biased picking from $\left(\xi_{(k),P_\mu}, k = 1, .., P_\mu\right)$. Its moment function is

$$\varphi_\eta(q) = \mathbb{E}\eta^q := \mathbb{E}\frac{1}{\theta}\sum_{k=1}^{P_\mu} \xi_{(k),P_\mu}^{q+1} = \frac{1}{\theta}\phi(q+1) \tag{5.4}$$

and η has probability distribution $F_\eta(x) = \frac{\mu}{\theta}\int_0^x zF_\xi(dz)$.

The waiting time paradox reads

$$\eta \succeq_{st} \xi,$$

since $F_\eta(x) \leq F_\xi(x)$ for all $x \in [0,1]$.

5.2 Spacings

As in the unbounded case, spacings of the Poisson partition of χ deserve interest.

Defining the spacings $\widetilde{\xi}_{(k),P_\mu} := \xi_{(k-1),P_\mu} - \xi_{(k),P_\mu}$ (with $\xi_{(0),P_\mu} := 1$ and $\xi_{(P_\mu+1),P_\mu} := 0$), $k = 1, .., P_\mu + 1$, then, $\left(\widetilde{\xi}_{(k),P_\mu}, k = 1, .., P_\mu + 1\right)$ constitutes a new sequence of $(0,1)$ −valued random variables with clearly $\sum_{k=1}^{P_\mu+1} \widetilde{\xi}_{(k),P_\mu} = 1$. Let for example

$$\widetilde{\phi}(\beta) := \mathbb{E}\sum_{k=1}^{P_\mu+1} \widetilde{\xi}_{(k),P_\mu}^{\beta}.$$

With $u_{(0),P_\mu} := 0$ and $u_{(P_\mu+1),P_\mu} := 1$, it can be obtained in general from

$$\widetilde{\phi}(\beta) = \mathbb{E}\sum_{k=1}^{P_\mu+1}\left[\left(\overline{F}^{-1}\left(u_{(k-1),P_\mu}\right) - \overline{F}^{-1}\left(u_{(k),P_\mu}\right)\right)^{\beta}\right] \tag{5.5}$$

so that, given $P_\mu = p$, the joint law of $\left(u_{(k-1),p}; u_{(k),p}\right)$ is needed to compute $\widetilde{\phi}(\beta)$ or the structural measure σ such that $\widetilde{\phi}(\beta) = \int_0^1 x^\beta \sigma(dx)$.

Example:

Let $\Pi(dx) = \mu \mathbf{I}_{x\in(0,1)}dx$. Then $F(dx)$ is the uniform distribution ($\overline{F}^{-1}(x) = 1-x$) and $\theta = \mu/2$. We obtain $\mathbb{E}\sum_{k=1}^{P_\mu}\xi_{(k),P_\mu}^{\beta} = \frac{2\theta}{\beta+1} = \phi(\beta)$. The random variable ξ is uniform and the size-biased picking random variable η has distribution beta$(2,1)$.

Concerning spacings, we have

$$\widetilde{\phi}(\beta) = \mathbb{E}\sum_{k=1}^{P_\mu+1}\left(\xi_{(k-1),P_\mu} - \xi_{(k),P_\mu}\right)^{\beta} = \int_0^1 x^\beta \sigma(dx).$$

The tail of the structural measure for spacings reads

$$\begin{aligned}\int_x^1 \sigma(dz) =: \overline{\sigma}(x) &= \mathbb{E}\sum_{k=1}^{P_\mu+1}\mathbb{P}\left(\xi_{(k-1),P_\mu} - \xi_{(k),P_\mu} > x\right)\\ &= \sum_{p\geq 0}\frac{e^{-\mu}\mu^p}{p!}(p+1)(1-x)^p = e^{-\mu x}(1+\mu(1-x))\end{aligned}$$

recalling that $\overline{F}_{\widetilde{\xi}_{(k),:p}}(x) = (1-x)^p$ is the tail distribution function of uniform spacings $\widetilde{\xi}_{(k),p} := \xi_{(k-1),p} - \xi_{(k),p}$ (with $\xi_{(0),p} := 1$ and $\xi_{(p+1),p} := 0$), for any $k = 1,..,p+1$. $\square$

References

[1] Derrida B, Statistical properties of randomly broken objects and of multivalley structures in disordered systems, 1987, *Journal of Physics A: Math. Gen.,* **20**, 5273-5288.

[2] Feller W, *An Introduction to Probability Theory and its Applications,* 1971, Vol. **2**, (Wiley, New York).

[3] Derrida B, Random energy model: an exactly solvable model of disordered systems, 1981, *Physical Review B.,* **24**, 2613-2626.

[4] Kingman J.F.C, *Poisson Processes*, 1993, Clarendon Press, Oxford.

[5] Tavaré S and Ewens W.J, Multivariate Ewens distribution. Chapter **41** in *Discrete Multivariate Distributions*, Edts N.L. Johnson, S. Kotz and N. Balakrishnan, 1997, p. 232-246, (Wiley, New York).

[6] Pitman J, Random discrete distributions invariant under size-biased permutation, 1996, *Advances in Applied Probability,* **28**, 525-539.

[7] Pitman J, Coalescents with multiple collisions. *Annals of Probability*, 1999, **27**, n° 4, 1870–1902.

[8] Pitman J, Poisson-Dirichlet and GEM invariant distributions for split-and-merge transformation of an interval partition, 2002, *Combinatorics, Probability and Computing*, **11**, n° 5, 501–514.

[9] Pitman J. and Yor M, The two parameter Poisson-Dirichlet distribution derived from a stable subordinator, 1997, *Annals of Probability*, **25**, 855-900.

[10] Kingman J.F.C, Random discrete distributions, 1975, *Journal of the Royal Statistical Society, Series B*, **37**, 1–22.

[11] Holst L, The Poisson-Dirichlet distribution and its relatives revisited, 2001, Preprint of the *Royal Institute of Technology*, Stockholm, Sweden.

[12] Bertoin J, *Lévy processes*, 1996, Cambridge University Press, Cambridge.

[13] Steutel F.W. and van Harn K, *Infinite Divisibility of Probability Distributions on the Real Line,* 2004, Pure and Appl. Math. vol. **259**, Marcel Dekker Pub., New-York, Basel.

[14] Ruelle D, Mathematical reformulation of Derrida's REM and GREM, 1987, *Communications in Mathematical Physics,* **108**, 225-239.

[15] Koukiou F. and Picco P, Poisson point processes, cascades, and random coverings of R^n, 1991, *Journal of Statistical Physics*, **62**, n° 1-2, 481–489.

[16] Neveu J, *Processus Ponctuels.* (École d'été de probabilités de Saint-Flour VI, 1976), 1978, Lecture Notes in Mathematics, vol. 598, Springer.

[17] Lindenberg K. and West B.J, The first, the biggest, and other such considerations, 1986, *Journal of Statistical Physics,* **42** n° 1/2, 201-243.

[18] Ignatov Ts, A constant arising in the asymptotic theory of symmetric groups and Poisson-Dirichlet measures, 1982, *Theory of Probability and its Applications*, **27**, n° 1, 136–147.

[19] Arratia R, Barbour A. D and Tavaré S, The Poisson-Dirichlet distribution and the scale-invariant Poisson process, 1999, *Combinatorics, Probability and Computing*, **8**, n°. 5, 407–416.

[20] Pitman J, *Poisson-Kingman Partitions,* 2003, Statistics and Science: a Festschrift for Terry Speed, 1–34, IMS Lecture Notes Monogr. Ser.,40, Inst. Math. Statist., Beachwood, OH.

[21] Perman M, Order statistics for jumps of normalized subordinators, 1993, *Stochastic Processes and Applications*, **46**, 267-281.

[22] Hirth U. M, A Poisson approximation for the Dirichlet law, the Ewens sampling formula and the Griffiths-Engen-McCloskey law by the Stein-Chen coupling method, 1997, *Bernoulli* **3**, n°. 2, 225–232.

[23] Hirth U. M, From GEM back to Dirichlet via Hoppe's urn, 1997, *Combinatorics, Probability and Computing,*. **6**, n° 2, 185–195.